Thin, A., Kotolevskii, N., Andrieu, C., Durmus, A., Moulines, E., & Panov, M. (2020). Nonreversible MCMC from conditional invertible transforms: a complete recipe with convergence guarantees. *arXiv*. https://arxiv.org/abs/2012.15550

Nonreversible MCMC from conditional invertible transforms: a complete recipe with convergence guarantees

Achille Thin[1], Nikita Kotolevskii[2], Christophe Andrieu[3], Alain Durmus[4]

Eric Moulines[15], Maxim Panov[2]

Abstract

Markov Chain Monte Carlo (MCMC) is a class of algorithms to sample complex and high-dimensional probability distributions. The Metropolis-Hastings (MH) algorithm, the workhorse of MCMC, provides a simple recipe to construct reversible Markov kernels. Reversibility is a tractable property which implies a less tractable but essential property here, invariance. Reversibility is however not necessarily desirable when considering performance. This has prompted recent interest in designing kernels breaking this property. At the same time, an active stream of research has focused on the design of novel versions of the MH kernel, some nonreversible, relying on the use of complex invertible deterministic transforms. While standard implementations of the MH kernel are well understood, aforementioned developments have not received the same systematic treatment to ensure their validity. This paper fills the gap by developing general tools to ensure that a class of nonreversible Markov kernels, possibly relying on complex transforms, has the desired invariance property and lead to convergent algorithms. This leads to a set of simple and practically verifiable conditions.

1 Introduction

Being able to simulate from a probability distribution, say π defined on a measurable space $(\mathsf{Z}, \mathcal{Z})$ and referred to as the target distribution hereafter, is a ubiquitous task. Markov chain Monte Carlo methods (MCMC) is an important body of versatile techniques to sample from π. They consist of simulating realisations of time-homogeneous Markov chains $(Z_k)_{k\in\mathbb{N}}$ of invariant distribution π which possess the property that their realised states can be used to mimic samples from π, that is $Z_k \sim \pi$ approximately, but with arbitrary precision, and approximate expectations with respect to π – more precise statements are provided in Theorem 1 and we refer to these, for now, lose concepts as "convergence". We denote by P the Markov kernel associated with $(Z_k)_{k\in\mathbb{N}}$.

[1]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.
achille.thin@polytechnique.edu, eric.moulines@polytechnique.edu
[2]CDISE, Skolkovo Institute of Science and Technology, Moscow, Russia
Nikita.Kotelevskii@skoltech.ru, M.Panov@skoltech.ru
[3]School of Mathematics, University of Bristol, UK. c.andrieu@bristol.ac.uk
[4]Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, F-91190 Gif-sur-Yvette, France.
alain.durmus@cmla.ens-cachan.fr
[5]CS Departement, HSE University, Russian Federation

Metropolis-Hastings (MH) is a popular strategy to design such a Markov kernel. In its most common form, the "textbook" MH kernel samples the $(k+1)$-th state Z_{k+1} of $(Z_k)_{k\in\mathbb{N}}$ as follows: (1) sample a proposal $Y_{k+1} \sim Q(Z_k, \cdot)$; (2) set $Z_{k+1} = Y_{k+1}$ with probability $\alpha(Z_k, Y_{k+1})$; otherwise, set $Z_{k+1} = Z_k$, where $Q\colon \mathsf{Z} \times \mathcal{Z} \to [0,1]$ is a Markov kernel and $\alpha\colon \mathsf{Z} \times \mathsf{Z} \to [0,1]$ is the acceptance probability. General conditions on π, Q and α in order to ensure invariance and convergence of $(Z_k)_{k\in\mathbb{N}}$ have been known for some time. In the particular situation where π and $\{Q(z,\cdot), z \in \mathsf{Z}\}$ have densities π and $\{q(z,\cdot), z \in \mathsf{Z}\}$ with respect to a common dominating measure and are positive everywhere one can choose $\alpha(z,z') = \min\{1, \pi(z')q(z',z)/[\pi(z)q(z,z')]\}$ and define a convergent algorithm.

Contribution #1: a complete recipe for $(\pi, S)-$reversible kernels. In the context of MCMC the $\pi-$invariance property of P is traditionally the consequence of a stronger property, $\pi-$reversibility (related to detailed balance Fang et al. (2014)), which is however more tractable in practice. The MH Markov kernel is designed to satisfy this property. However, there has been a re-kindled interest in the development of "nonreversible" algorithms Turitsyn et al. (2011); Hukushima & Sakai (2013); Ma et al. (2016); Ottobre (2016); Bierkens & Roberts (2017); Neklyudov et al. (2020); Sherlock & Thiery (2019); Gustafson (1998) which come with the promise of removing the backtracking behaviour of reversible algorithms, and hence speed-up convergence. Our first contribution (Section 2) is **(a)** a review of $(\pi, S)-$reversibility, related to the modified detailed balance condition Fang et al. (2014), a generalisation of reversibility behind most so-called "nonreversible" MCMC algorithms and **(b)** a method generalizing the MH rule to obtain (π, S)-reversible kernels from arbitrary proposal kernels Q. This is a generalisation of Tierney (1998) which aims to provide a unifying and firm theoretical footing to recent and future contributions. The framework encompasses, for example, both the scenarios where π and Q have common dominating measure or when Q corresponds to a deterministic mapping.

New challenges. Novel applications have led to the development of highly sophisticated extensions of this basic scheme, prompted in particular by recent developments in the context of probability density representation with normalising flows Baptista et al. (2020); Prangle (2019); Papamakarios et al. (2019), invertible neural networks Ardizzone et al. (2019). For example, following the realisation that the textbook MH can be generalised by combining deterministic invertible mappings of the current state and a source of randomness in the proposal stage, some authors have proposed using complex mappings involving both non-linearities and the composition of multiple layers Albergo et al. (2019); Thin et al. (2020); Spanbauer et al. (2020), while Sherlock & Thiery (2019); Gustafson (1998) explore the use of nonreversible Markov kernels. However it is not always clear that the resulting algorithms are convergent. In particular application of Markov chain theory may seem difficult at first sight given the new levels of complexity involved. Our aim in this paper is to provide users with simple to use theoretical guarantees ensuring validity of the algorithms.

Contribution #2: easy ready-made convergence results. Proving convergence of MH methods can be delicate in general. However, in the π-reversible case, Mengersen & Tweedie (1996) and Tierney (1994) have derived simple conditions ensuring convergence of P in the case where π and Q share a common dominating measure μ, for example the Lebesgue measure when $\mathsf{Z} = \mathbb{R}^d$.

Theorem 1 (Convergence of textbook MH). *Assume that π is not a Dirac mass function and has common $\sigma-$finite dominating measure μ with $\{Q(z,\cdot), z \in \mathsf{Z}\}$. Denote π and $\{q(z,\cdot), z \in \mathsf{Z}\}$ the resulting densities. Suppose in addition that π is not a Dirac mass and $Q(z, \mathsf{Z}^+) = 1$ for any $z \notin \mathsf{Z}^+$ with $\mathsf{Z}^+ = \{z \in \mathsf{Z} : \pi(z) > 0\}$. If for any $z' \in \mathsf{Z}$ such that $\pi(z') > 0$ we have $q(z,z') > 0$ for any*

$z \in \mathsf{Z}$, then for any $f \colon \mathsf{Z} \to \mathbb{R}$ such that $\pi(|f|) < \infty$, almost surely it holds that

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f(Z_i) = \pi(f) \ . \tag{1}$$

In addition, for all $z \in \mathsf{Z}$

$$\lim_{n\to\infty} \|P^n(z,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} = 0 \ . \tag{2}$$

Our second contribution is to provide similarly simple and easy to use conditions to establish convergence for (π, S)-reversible kernel. We translate these conditions to cover complex proposal mechanisms based on conditional invertible neural transform ensuring that basic convergence properties hold in these novel settings. As we shall see some of the results lead to simple implementation suggestions ensuring that conclusions similar to those of Theorem 1 hold. Establishing these properties is often overlooked and a necessary prerequisite to any more refined analysis characterising their performance, such as quantitative finite time convergence bounds.

Contribution #3: application to particular MCMC algorithms. We show how our conditions and construction can be used in practice to design (π, S)-reversible kernels which come with convergence guarantees. We first work out a generalization of the Hamiltonian Monte Carlo algorithm in which the gradients of the log-density in the leap frog steps are replaced by general neural transforms Neal (2011); Sohl-Dickstein et al. (2014). Next, we derive and analyse two lifted Markov kernels Diaconis et al. (2000); Chen et al. (1999); Turitsyn et al. (2011); Neklyudov et al. (2020) covering obtained using conditional invertible transforms on an augmented space. Our experimental results (postponed to Supplementary paper) show numerically the benefits of nonreversibility in several sampling experiments.

The proofs of the main results and some facts, followed by a $*$, can be found in the supplementary material and, for example, (S#) refers to the #-th equation in the supplement. The standard notation and definitions used are precisely described in the supplementary Appendix A for the reader's convenience.

2 (π, S)-reversibility and the Generalized MH rule

There has recently been renewed interest in the design of π-invariant Markov kernels which are nonreversible. In many scenarios, departing from reversibility can both improve the mixing time and reduce the asymptotic variance of resulting estimators. It has been shown in Andrieu & Livingstone (2019) that these nonreversible Markov kernels fall under the same common framework of (π, S)-reversibility (introduced below) which encompasses the modified (or skew) detailed balance conditions. Before proceeding further, additional notations are needed. Let s be an involution on Z, $s \circ s = \mathrm{Id}$ and S be the associated kernel $S(z, A) = 1_A\big(s(z)\big)$, $z \in \mathsf{Z}$, $A \in \mathcal{Z}$. Let $\check{\mu}$ be a σ-finite measure on the product space $(\mathsf{Z}^2, \mathcal{Z}^{\otimes 2})$ (the diacritic $\check{}$ is used to denote measures on the product space Z^2). Denote by $\check{\mu}^s = (F_s)_\# \check{\mu}$ the pushforward of $\check{\mu}$ by the transform $F_s(z, z') = \big(s(z'), s(z)\big)$:

$$\check{\mu}^s(C) = \int 1_C\big(s(z'), s(z)\big)\check{\mu}\big(\mathrm{d}(z, z')\big) \ , C \in \mathcal{Z}^{\otimes 2} \ .$$

Note that F_s is an involution $F_s \circ F_s = \mathrm{Id}$ which implies that $(\check{\mu}^s)^s = \check{\mu}$.

Definition 2 (after Andrieu & Livingstone (2019))**.** *The measure $\check{\mu}$ is s-symmetric if $\check{\mu} = \check{\mu}^s$. The sub-Markovian kernel P is (π, S)-reversible if the measure $\check{\mu}_P$ defined as $\check{\mu}_P\big(\mathrm{d}(z,z')\big) = \pi(\mathrm{d}z)P(z, \mathrm{d}z')$ is s-symmetric.*

It is established in Appendix C.1 that P is (π, S)-reversible if it satisfies the **skew detailed balance** condition,

$$\pi(\mathrm{d}z)P(z, \mathrm{d}z') = s_{\#}\pi(\mathrm{d}z')SPS(z', \mathrm{d}z) . \tag{3}$$

In particular, if $s_{\#}\pi = \pi$ and P is a Markov kernel, then π is invariant for P. We assume that the condition $s_{\#}\pi = \pi$ is in force in the rest of the paper. Note that for $s = \mathrm{Id}$ we recover the standard detailed balance condition (see Appendix B).

2.1 Generalized Metropolis-Hastings

The MH algorithm gives a method to transform any proposal Markov kernel Q into a π-reversible Markov kernel. We derive a Generalized Metropolis-Hastings (GMH) rule to turn Q into a (π, S)-reversible Markov kernel. We then apply this condition to the case where $\pi(\mathrm{d}z')$ and $Q(z, \mathrm{d}z')$ have a density w.r.t. to a common dominating measure, and to the case where $Q(z, \mathrm{d}z') = \delta_{\Phi(z)}(\mathrm{d}z')$ for $\Phi\colon \mathsf{Z} \to \mathsf{Z}$. We first establish a simple necessary and sufficient condition on the proposal kernel Q and the acceptance probability function $\alpha\colon \mathsf{Z}^2 \to [0, 1]$ for the resulting (sub-Markovian) kernel

$$Q_\alpha(z, \mathrm{d}z') := \alpha(z, z')Q(z, \mathrm{d}z') , \tag{4}$$

to be (π, S)-reversible. A subset $A \subset \mathsf{Z}^2$ is said to be s-symmetric if $(z, z') \in A$ if and only if $\big(s(z'), s(z)\big) \in A$. We denote

$$\check{\nu}\big(\mathrm{d}(z, z')\big) := \pi(\mathrm{d}z)Q(z, \mathrm{d}z') .$$

The following result provides us with a key instrument to work with the densities of $\check{\nu}$ and its pushforward $\check{\nu}^s$ in full generality.

Proposition 3. *Set $\check{\lambda} = \check{\nu} + \check{\nu}^s$, $h = \mathrm{d}\check{\nu}/\mathrm{d}\check{\lambda}$ and $A_{\check{\nu}} = \{h \times h \circ F_s > 0\} \in \mathcal{Z}^{\otimes 2}$. Then, the restrictions $\check{\nu}_A(\cdot) = \check{\nu}(\cdot \cap A_{\check{\nu}})$ and $\check{\nu}^s_A(\cdot) = \check{\nu}^s(\cdot \cap A_{\check{\nu}})$ are equivalent and $\check{\nu}_{A,\mathrm{c}}(\cdot) = \check{\nu}(\cdot \cap A^{\mathrm{c}}_{\check{\nu}})$ and $\check{\nu}^s_{A,\mathrm{c}}(\cdot) = \check{\nu}^s(\cdot \cap A^{\mathrm{c}}_{\check{\nu}})$ are mutually singular. In addition, define, for $(z, z') \in A_{\check{\nu}}$, $r(z, z') = h(z, z')/h\big(s(z'), s(z)\big)$. Then, r is a version of the density of $\check{\nu}_A$ w.r.t. $\check{\nu}^s_A$, i.e. $r = \mathrm{d}\check{\nu}_A/\mathrm{d}\check{\nu}^s_A$ and $r(z, z') = 1/r \circ F_s(z, z')$ for all $(z, z') \in A_{\check{\nu}}$.*

The following result applies Proposition 3 and extends the seminal result (Tierney, 1998, Theorem 2) to the (π, S)-reversible case.

Theorem 4. *The sub-Markovian kernel Q_α in (4) is (π, S)-reversible if and only if the following conditions hold.*
(i) The function α is zero $\check{\nu}$-a.e.on $A^{\mathrm{c}}_{\check{\nu}}$.
(ii) The function α satisfies $\alpha(z, z')r(z, z') = \alpha(s(z'), s(z))$ $\check{\nu}$-a.e.on $A_{\check{\nu}}$.

Similarly to the π-reversible case, we define the generalized Metropolis-Hastings (GMH) rejection probability by

$$\alpha(z, z') = \begin{cases} \mathsf{a}\left(\frac{h(s(z'),s(z))}{h(z,z')}\right) & h(z, z') \neq 0, \\ 1 & h(z, z') = 0, \end{cases} \tag{5}$$

where $\mathsf{a}\colon \mathbb{R}_+^* \to [0,1]$ satisfies $\mathsf{a}(0) = 0$ and for $t \in \mathbb{R}_+^*$,

$$t\mathsf{a}(1/t) = \mathsf{a}(t)\ . \tag{6}$$

Then α satisfies the conditions (i)-(ii) of Theorem 4, see Appendix C.4. We may take for example $\mathsf{a}(t) = \min(1,t)$ or $\mathsf{a}(t) = t/(1+t)$ which correspond to the classical Metropolis-Hastings and Barker ratio respectively.

We can obtain the GMH Markov kernel P which is (π, S)-reversible by adding Dirac masses:

$$P(z, \mathrm{d}z') = Q_\alpha(z, \mathrm{d}z') + a(z)\delta_z(\mathrm{d}z') + b(z)\delta_{s(z)}(\mathrm{d}z') \tag{7}$$

with a, b nonnegative, measurable satisfying $a(z) = a(s(z))$ and $a(z) + b(z) = 1 - Q_\alpha(z, \mathsf{Z})$; see Appendix C.5. In the sequel, we focus on the case $a(z) = 0$ and $b(z) = 1 - Q_\alpha(z, \mathsf{Z})$.

2.2 GMH for particular proposal maps

We now specialize (5) to the case where π and Q admit a common dominating measure and the case where Q is deterministic.

Proposal with densities. Suppose there is a common dominating measure μ on $(\mathsf{Z}, \mathcal{Z})$ such that $\pi(\mathrm{d}z) = \pi(z)\mu(\mathrm{d}z)$, $Q(z, \mathrm{d}z') = q(z, z')\mu(\mathrm{d}z')$ and that μ is invariant by s, *i.e.* $s_\#\mu = \mu$. In this scenario, we have (see Appendix C.6)

$$A_{\tilde{\nu}} = \left\{\pi(z)q(z, z') \times \pi(z')q\big(s(z'), s(z)\big) > 0\right\}\ . \tag{8}$$

In addition, we obtain using (5) that

$$\alpha(z, z') = \begin{cases} \mathsf{a}\left[\frac{\pi(z')q(s(z'), s(z))}{\pi(z)q(z, z')}\right] & \pi(z)q(z, z') \neq 0, \\ 1 & \pi(z)q(z, z') = 0\ . \end{cases} \tag{9}$$

Theorem 1 exploits the fact that in the π-reversible scenario the MH kernel is π-irreducible if the condition $\pi(z') > 0$ implies that $q(z, z') > 0$ Mengersen & Tweedie (1996). This result can be extended to the (π, S)-reversible case as follows.

Lemma 5. *The GMH Markov kernel P in (7) is π-irreducible if, $\pi(z') > 0$ implies that, for all $z \in \mathsf{Z}$, $q(z, z') > 0$ and $q\big(s(z), s(z')\big) > 0$.*

In the π-reversible case, (Tierney, 1994, Corollary 2) shows that the π-irreducibility condition implies that the GMH Markov kernel P (7) is Harris recurrent and aperiodic. These two properties have consequences that are very important in practice: the convergence in total variation of the iterates of the kernel to the invariant distribution and the ergodic theorem become valid also for all the initial conditions. These results extend to (π, S)-Markov kernels (see Appendix C.7).

Theorem 6. *Let P be defined as in (7), with $a(z) = 0$ and $b(z) = 1 - Q_\alpha(z, \mathsf{Z})$. Assume that for any $z' \in \mathsf{Z}$, $\pi(z') > 0$ implies $q(z, z') \times q\big(s(z), s(z')\big) > 0$ for any $z \in \mathsf{Z}$. Suppose in addition that π is not a Dirac mass and $Q(z, \mathsf{Z}^+) = 1$ for any $z \notin \mathsf{Z}^+$ with $\mathsf{Z}^+ = \{z \in \mathsf{Z} :\ \pi(z) > 0\}$. The conclusions of Theorem 1 hold.*

Deterministic proposal. Suppose now that Φ is a one-to-one mapping from Z onto Z such that

$$\Phi^{-1} = s \circ \Phi \circ s \,. \tag{10}$$

We consider the deterministic proposal kernel $Q(z, \mathrm{d}z') = \updelta_{\Phi(z)}(\mathrm{d}z')$: when the current state is z, the proposal is $\Phi(z)$. Condition (10) implies that $F = s \circ \Phi$ is an involution. Our setting covers involutive MCMC – corresponding to the case $s = \mathrm{Id}$ introduced in (Tierney, 1998, Section 2) and more recently in Neklyudov et al. (2020).

In this scenario we have that (see Appendix C.8) $\check{\nu}\big(\mathrm{d}(z, z')\big) = \pi(\mathrm{d}z)\updelta_{\Phi(z)}(\mathrm{d}z')$ and $\check{\nu}^s\big(\mathrm{d}(z, z')\big) = \pi(\mathrm{d}z')\updelta_{\Phi^{-1}(z')}(\mathrm{d}z)$. The function h defined in Proposition 3 is given by $h(z, z') = 1_{\Phi(z)}(z')k(z)$ with

$$k(z) = \frac{\mathrm{d}\pi}{\mathrm{d}\lambda}(z), \quad \lambda = \pi + (\Phi^{-1})_{\#}\pi \,. \tag{11}$$

Theorem 4 is satisfied with the acceptance probability given by $\alpha\big(z, \Phi(z)\big) = \bar{\alpha}(z)$ with

$$\bar{\alpha}(z) = \mathsf{a}\left(\frac{k\big(s \circ \Phi(z)\big)}{k(z)}\right) \tag{12}$$

if $k(z) > 0$ and $\bar{\alpha}(z) = 1$, otherwise. Of course, there is no need to define $\alpha(z, z')$ for $z' \neq \Phi(z)$. A special case of interest is when $\mathsf{Z} = \mathbb{R}^d$ and the target distribution $\pi(\mathrm{d}z) = \pi(z)\mathrm{Leb}_d(\mathrm{d}z)$ has a density w.r.t. the Lebesgue measure on $\mathbb{R}^d$. Here the dominating measure λ is given by

$$\lambda(\mathrm{d}z) = \{\pi(z) + \pi \circ \Phi(z)\, \mathrm{J}_{\Phi}(z)\}\mathrm{Leb}_d(\mathrm{d}z) \,,$$

where J_f denotes the absolute value of the Jacobian determinant of f. Then, the density $k(z)$ is given by

$$k(z) = \frac{\pi(z)}{\pi(z) + \pi \circ \Phi(z)\, \mathrm{J}_{\Phi}(z)}$$

and the acceptance ratio $\bar{\alpha}(z)$ takes the simple form

$$\bar{\alpha}(z) = \mathsf{a}\left(\frac{\pi \circ \Phi(z)\, \mathrm{J}_{\Phi}(z)}{\pi(z)}\right) \tag{13}$$

if $\pi(z) \neq 0$ and $\bar{\alpha}(z) = 1$ otherwise (see Appendix C.9). We obtain the same acceptance ratio given by (Neklyudov et al., 2020, Eq. (5)), which derive this expression in the special case $s = \mathrm{Id}$ and thus $\Phi^{-1} = \Phi$ is an involution. It is perhaps striking that the acceptance ratio **does not depend** on s: this comes from the fact the target distribution π is invariant by s. This setting encompasses many algorithms, HMC Duane et al. (1987); Neal (2011), and NICE-MC Song et al. (2017) – see below, Neklyudov et al. (2020) and the references therein. Of course, in most cases, (π, S)-reversible deterministic Markov kernels are not π-irreducible and Harris recurrent. They can nevertheless be important building blocks of Markov kernels as in the HMC construction.

3 Applications and examples

3.1 Generalized Hamiltonian Dynamics

We first consider generalizations of the Hamiltonian Monte Carlo algorithm (see Neal (2011); Sohl-Dickstein et al. (2014)). These methods might also be seen as a special case of NICE (Non-linear Independent Components Estimation) MCMC methods Song et al. (2017); Neklyudov et al.

(2020). The objective is to sample a distribution on $\mathbb{R}^d$ of density π_0 w.r.t. the Lebesgue measure. We use a data augmentation approach which consists of adding a "momentum" variable with stationary distribution admitting a symmetric density φ on $\mathbb{R}^d$ w.r.t. the Lebesgue measure, e.g. $\varphi(-p) = \varphi(p)$. More precisely, on the extended state space $\mathsf{Z} = \mathbb{R}^{2d}$, we consider the extended target density defined by $\pi(x, p) = \pi_0(x)\varphi(p)$ and the Markov chain $\big(Z_i = (X_i, P_i)\big)_{i\in\mathbb{N}}$. The involution is taken to be $s(x, p) = (x, -p)$. By construction, $s_{\#}\pi = \pi$.

We first show how to construct a (π, S)-reversible Markov kernel on $\mathbb{R}^{2d}$ using modified leapfrog integrators. Let $m \in \mathbb{N}$ and $\{M_i, N_i\}_{i=1}^m$ be C^1 functions on $\mathbb{R}^d$. We define a mapping $\Phi(x, p) = \mathrm{F}_m \circ \cdots \circ \mathrm{F}_1(x, p)$ on $\mathbb{R}^{2d}$ where F_i is given by $(x_{i+1}, p_{i+1}) = \mathrm{F}_i(x_i, p_i)$ where for $h > 0$

$$\begin{cases} p_{i+1/2} &= p_i + hM_i(x_i) \ , \\ x_{i+1} &= x_i + hp_{i+1/2} \ , \\ p_{i+1} &= p_{i+1/2} + hN_i(x_{i+1}) \ . \end{cases} \tag{14}$$

It is easily seen that F_i is a C^1 diffeomorphism on $\mathbb{R}^{2d}$ to $\mathbb{R}^{2d}$ with $\mathrm{J}_{\mathrm{F}_i}(x, p) = 1$. Moreover, if for any $i \in \{1, \dots, m\}$, $M_i = N_{m+1-i}$, then $s \circ \Phi \circ s = \Phi^{-1}$; see Appendix D.1. We assume in the sequel that this condition holds. Consider now the Markov kernel

$$\begin{aligned} P((x, p), \mathrm{d}(y, q)) &= \bar{\alpha}(x, p)\updelta_{\Phi(x,p)}(\mathrm{d}(y, q)) \\ &\quad + (1 - \bar{\alpha}(x, p))\updelta_{(x,-p)}(\mathrm{d}(y, q)) \ , \end{aligned} \tag{15}$$

$$\text{with} \qquad \bar{\alpha}(x, p) = \mathsf{a}\left(\pi \circ \Phi(x, p)/\pi(x, p)\right) \ , \tag{16}$$

if $\pi(x, p) > 0$ and $\bar{\alpha}$ is equal to 1 otherwise. Using (13), P is (π, S)-reversible, but is deterministic and therefore not ergodic. A standard approach to address this issue, used in the context of HMC algorithms, is to refresh the momentum between two successive moves according to a Markov transition preserving the distribution φ. A particular choice consists of sampling the velocity afresh from φ before applying the kernel (39). More precisely, we define the Markov chain $(X_i)_{i\in\mathbb{N}}$ by the following recursion. From a state X_k, the $k+1$-th iterate is defined by: 1. sample P_{k+1} from φ and set $(Y_{k+1}, Q_{k+1}) = \Phi(X_k, P_{k+1})$; accept $X_{k+1} = Y_{k+1}$ with probability $\bar{\alpha}(X_k, P_{k+1})$ and reject $X_{k+1} = X_k$ otherwise. In this case one can check that $(X_i)_{i\in\mathbb{N}}$ is a Markov chain on $\mathbb{R}^d$ of kernel, obtained by marginalisation of (7) w.r.t. the momentum distribution,

$$K(x, \mathrm{d}y) = K_\alpha(x, \mathrm{d}y) + \{1 - \bar{\alpha}(x)\}\updelta_x(\mathrm{d}y) \ , \tag{17}$$

where $\bar{\alpha}(x) = K_\alpha(x, \mathbb{R}^d)$ and denoting $G_x(p) = \mathrm{proj}_1 \circ \Phi(x, p)$, $\mathrm{proj}_1(x, p) = x$,

$$K_\alpha(x, \mathrm{d}y) = \int \bar{\alpha}(x, p)\varphi(p)\updelta_{G_x(p)}(\mathrm{d}y)\mathrm{d}p$$

If for any $x \in \mathbb{R}^d$, $p \mapsto G_x(p)$ is a diffeomorphism on $\mathbb{R}^d$, then Theorem 6 can be applied. In such case, $K_\alpha(x, \mathrm{d}y) = \alpha(x, y)q(x, y)$ with

$$\alpha(x, y) = \mathsf{a}\left(\frac{\pi_0(y)\varphi\big\{H_x\big(G_x^{-1}(y)\big)\big\}}{\pi_0(x)\varphi\big(G_x^{-1}(y)\big)}\right) \ , \tag{18}$$

$$q(x, y) = \varphi\big(G_x^{-1}(y)\big)\,\mathrm{J}_{G_x^{-1}}(y) \ , \tag{19}$$

and $H_x(p) = \mathrm{proj}_2 \circ \Phi(x, p)$ and $\mathrm{proj}_2(x, p) = p$. The expression of $\alpha(x, y)$ is only of theoretical interest and is not needed to implement the algorithm. Of course, requiring that G_x is a diffeomorphism imposes conditions on F_i, $i \in \{1, \dots, m\}$ and Theorem 24 (see Appendix D.1.2).

Theorem 7. *Assume that $\varphi > 0$, $\varphi(-p) = \varphi(p)$ for all $p \in \mathbb{R}^d$ and for any $i \in \{1, \dots, m\}$, M_i and N_i are* L*-Lipschitz and $h \leqslant c_0/[\mathrm{L}^{1/2} m]$, where $c_0 \approx 0.3$ (see Theorem 24). Then for any $x \in \mathbb{R}^d$, $p \mapsto G_x(p)$ is a C^1-diffeomorphism.*

The proof of this result is along the same lines as the proof of (Durmus et al., 2017, Theorem 1) which focuses on the standard HMC algorithm.

A by-product of the proof of Theorem 7, is that, perhaps surprisingly (see (D.1.3))

$$q(y, x)/q(x, y) = \varphi\big(H_x \circ G_x^{-1}(y)\big)/\varphi\big(G_x^{-1}(y)\big) , \tag{20}$$

implying that α (18) is the textbook MH acceptance ratio corresponding to q in (19), and the Markov kernel K (17) is therefore π-reversible. It easily checked that this kernel satisfies the conditions of Theorem 1 and the convergence results apply.

The π_0-reversibility of K has the disadvantage of loosing the potentially advantageous non-backtracking (or persistency) features of P. It is possible to recover persistency by considering the mixture of kernels on the extended space $\mathbb{R}^{2d}$ $\omega P + (1-\omega)L$ where P is the deterministic kernel (39) and $L((x,p), \mathrm{d}(y,q)) = K(x, \mathrm{d}y)\varphi(q)\mathrm{d}q$. In words, we refresh independently the position and the momentum. The amount of persistency is controlled by ω. Theorem 7 establishes π_0-irreducibility of K (see its proof), which immediately implies π-irreducibility of L; see Appendix D.1.1 for a more detailed discussion.

3.2 Lifted kernels

In this section, we apply the results of Section 2 to lifted kernels introduced in Diaconis et al. (2000); Chen et al. (1999); Turitsyn et al. (2011); Michel (2016). As above, let π_0 be a target probability density on $\mathbb{R}^d$ w.r.t. the Lebesgue measure. We extend the state space with a direction, *i.e.* we consider $\mathsf{Z} = \mathbb{R}^d \times \mathsf{V}$ with $\mathsf{V} = \{-1, 1\}$ and the extended target distribution $\pi = \pi_0 \otimes [\{\delta_{-1} + \delta_1\}/2]$. In this scenario the involution is $s(x, v) = (x, -v)$.

Proposal with densities. Let $q_{-1}(x, \cdot), q_1(x, \cdot)$ be two transition densities w.r.t. the Lebesgue measure on $\mathbb{R}^d$. Consider a proposal kernel $Q((x,v), \mathrm{d}(y,w))$ with density $q((x,v),(y,w))$ with respect to $\mathrm{Leb}_d(\mathrm{d}y) \otimes \{\delta_{-1}(\mathrm{d}w) + \delta_1(\mathrm{d}w)\}$ given by

$$q\big((x,v),(y,w)\big) = \big\{\rho 1_v(w) + (1-\rho) 1_{-v}(w)\big\} q_w(x,y), \tag{21}$$

where $\rho \in (0,1)$. In words, starting from (x,v), we either "keep" $w = v$ with probability ρ or "flip" $w = -v$ the direction otherwise, and then propose a candidate y according to $q_w(x, \cdot)$. In the original implementation of the lifting procedure Turitsyn et al. (2011), ρ is set to 1; taking $\rho < 1$ simply prevents the algorithm from getting "stuck" in one direction which could impede convergence of the algorithm.

From (5) and (9), the acceptance ratio α writes, for $q_w(x,y)\pi_0(x) \neq 0$, see Appendix D.2,

$$\alpha\big((x,v),(y,w)\big) = \mathsf{a}\left(\frac{q_{-w}(y,x)\pi_0(y)}{q_w(x,y)\pi_0(x)}\right) , \tag{22}$$

and $\alpha\big((x,v),(y,w)\big) = 1$ otherwise, where a satisfies (6). The GMH kernel is given by (7) with $a(z) = 0$ and $b(z) = 1 - Q_\alpha\big((x,v), \mathsf{Z}\big)$. Note that if the proposal move is rejected, then the direction is automatically flipped.

In the case $q_{-1} = q_1$, then the acceptance probability α (22) does not depend on v, w and the GMH kernel (7) can be marginalized w.r.t. v yielding the π_0-reversible MH algorithm of proposal density q_1. Since $\rho \in (0,1)$, the expression for q in (45) implies the following result.

Proposition 8. *Assume that for any $y \in \mathbb{R}^d$, $\pi_0(y) > 0$ implies $q_{-1}(x, y) > 0$ and $q_1(x, y) > 0$, for all $x \in \mathbb{R}^d$. Then the conditions of Theorem 6 hold, and the GMH kernel (7) is ergodic.*

Similarly to Section 3.1, the proposal densities $q_v(x, \cdot)$ are often associated to C^1-diffeomorphisms $G_{v,x} \colon p \mapsto G_{v,x}(p)$. From a state X_k, we sample P_{k+1} from φ positive density on $\mathbb{R}^d$ and set $Y_{k+1} = G_{V_k,X_k}(P_{k+1})$. In this case,

$$q_v(x, y) = \varphi\big(G_{v,x}^{-1}(y)\big)\, \mathrm{J}_{G_{v,x}^{-1}}(y) \; . \tag{23}$$

We illustrate the construction above with two examples of mappings G satisfying the conditions we consider.

Example 9 ((MALA-cIT) lifted kernel)**.** *Assume that π_0 is positive and continuously differentiable. For $x \in \mathbb{R}^d$, we define two transforms $G_{1,x}, G_{-1,x}$. For $G_{1,x}$, we set*

$$G_{1,x} \colon p \mapsto x + \gamma \nabla \log \pi(x) + \sqrt{2\gamma} p \; ,$$

which corresponds to the proposal of the Metropolis Adjusted Langevin Algorithm (MALA). In particular, for any $x \in \mathbb{R}^d$, the transformation $G_{1,x}$ is a C^1-diffeomorphism, with $\mathrm{J}_{G_{1,x}}(p) = (2\gamma)^{d/2}$ and

$$G_{1,x}^{-1}(y) = \{y - x - \gamma \nabla \log \pi(x)\}/\sqrt{2\gamma} \; .$$

For $G_{-1,x}$ we consider conditional invertible transforms Ardizzone et al. (2019)

$$G_{-1,x}(p) = \mathrm{G}_{K,x} \circ \cdots \circ \mathrm{G}_{1,x}(p) \; ,$$

where for $i \in \{1, \ldots, K\}$, $\mathrm{G}_{i,x}$ splits its input into two parts $(p_{i,1}, p_{i,2}) \in \mathbb{R}^{d_{i,1}} \times \mathbb{R}^{d-d_{i,1}}$ and applies affine transformations between them

$$\begin{aligned} p_{i+1,1} &= p_{i,1} \odot \exp\big(R_{i,1}(p_{i,2}, x)\big) + M_{i,1}(p_{i,2}, x) \; , \\ p_{i+1,2} &= p_{i,2} \odot \exp\big(R_{i,2}(p_{i+1,1}, x)\big) + M_{i,2}(p_{i+1,1}, x) \; . \end{aligned}$$

Here $R_{i,1}, M_{i,1}$ (resp. $R_{i,2}, M_{i,2}$) are any functions from $\mathbb{R}^{d_{i,1}}$ (resp. $\mathbb{R}^{d-d_{i,1}}$) to $\mathbb{R}^d$. This structure is an extension of the affine coupling block architecture suggested in Dinh et al. (2017). Note that for any $i \in \{1, \ldots, K\}$, $\mathrm{G}_{i,x}$ is a C^1-diffeomorphism on $\mathbb{R}^d$ of Jacobian determinant given by $\mathrm{J}_{\mathrm{G}_{i,x}}(p) = \exp\big(R_{i,1}(p_2, x) + R_{i,2}(p_1', x)\big)$. Therefore, $G_{-1,x}$ is a C^1-diffeomorphism with Jacobian determinant which can be explicitly computed. (23) gives a nonreversible MH algorithm with convergence guarantees provided by Proposition 8; see details in Appendix D.3.

A specific case corresponds to the choice

$$G_{v,x}(p) = \mathrm{proj}_1 \circ \Psi^v(x, p) \; ,$$

where Ψ is a C^1-diffeomorphism on $\mathbb{R}^{2d}$. We establish in the following result an alternative expression for α using (22) and (23), which relies on Ψ^v and J_{Ψ^v} and for which $\mathrm{J}_{G_{v,x}}$ is not required anymore (see Appendix D.4).

Lemma 10. *Assume that, for any $(x, v) \in \mathsf{Z}$, the mapping $G_{v,x}$ is a C^1-diffeomorphism on $\mathbb{R}^d$. Then, for any $x, y \in \mathbb{R}^d$, $v, w \in \mathsf{V}$, the acceptance ratio α defined in (22)is given by*

$$\mathsf{a}\left(\frac{\mu\Big(\Psi^w\big(x, G_{w,x}^{-1}(y)\big)\Big)}{\mu\big(x, G_{w,x}^{-1}(y)\big)}\, \mathrm{J}_{\Psi^w}\big(x, G_{w,x}^{-1}(y)\big) \right) \; ,$$

where $\mu(x, p) = \pi_0(x)\varphi(p)$.

This result is of practical interest because in many cases, the computation of $\mathrm{J}_{\Psi^v}(x,p)$ is much simpler than that of $\mathrm{J}_{G_{v,x}}(p)$. As an example, if Ψ is the generalized HMC transform $\Psi = \mathrm{F}_m \circ \cdots \circ \mathrm{F}_1$ where F_i is defined in (14), $\mathrm{J}_{\Psi^v}(x,p) = 1$ while $\mathrm{J}_{G_{v,x}}(p)$ has no simple closed-form expression.

Deterministic proposals. Using a C^1-diffeomorphism Ψ on $\mathbb{R}^{2d}$, we may also consider deterministic moves like in Section 3.1. Consider the extended state space $\mathsf{Z} = \mathbb{R}^{2d} \times \mathsf{V}$, the target distribution $\pi = \pi_0 \otimes \varphi \otimes [\{\updelta_{-1} + \updelta_1\}/2]$, where φ is a symmetric density w.r.t. Leb_d, and the involution $s(x,p,v) = (x,p,-v)$. Define $\Phi\big(x,p,v\big) = \big(\Psi^v(x,p), v\big)$. Then, it is immediate to see that $s \circ \Phi \circ s = \Phi^{-1}$. We consider the deterministic proposal kernel

$$Q\big((x,p,v), \mathrm{d}(y,q,w)\big) = \updelta_{\Psi^v(x,p)}(\mathrm{d}(y,q))\updelta_v(\mathrm{d}w) \ .$$

In the case, the acceptance ratio (12) reads for $x, p \in \mathbb{R}^d$, $v \in \mathsf{V}$ satisfying $\pi_0(x)\varphi(p) > 0$

$$\bar{\alpha}(x,p,v) = \mathsf{a}\left(\mu\big(\Psi^v(x,p)\big)\, \mathrm{J}_{\Psi^v}(x,p)/\mu(x,p)\right) \ , \tag{24}$$

and is equal to 1 if $\mu(x,p) = 0$, where $\mu(x,p) = \pi_0(x)\varphi(p)$; see Appendix D.5.

Example 11 (L2HMC)**.** *Assume that π_0 is positive and continuously differentiable. Using the framework depicted above, we show how the L2HMC algorithm Levy et al. (2017) (Learning To Hamiltonian Monte Carlo) can be turned into a nonreversible MCMC method by considering the map*

$$\Psi(x,p) = G_K \circ \cdots \circ G_1(x,p) \ , \tag{25}$$

where $G_i = H_i \circ F_i \circ H_{i-1/2}$ with, for $\delta > 0$,

- *for $j \in \{i, i-1/2\}$, $H_j(x,p) = (x, H_{j,x}(p))$ with*

$$H_{j,x}(p) = p \odot \exp\left(\delta R_j^H(x)\right) + \delta\big[\nabla \log \pi_0(x) \odot \exp\left(\delta R_j^H(x)\right) + M_j^H(x)\big] \ .$$

Note that H_j is a C^1-diffeomorphism on $\mathbb{R}^{2d}$ of Jacobian $\mathrm{J}_{H_j}(x,p) = \exp\big(\delta R_j^H(x)\big)$.

- *$F_i(x,p) = \big(F_{i,p}(x), p\big)$, where $F_{i,p}$ splits its input into two parts x_1, x_2 and applies affine transformations*

$$x_1' = x_1 \odot \exp\big(\delta R_{i,1}^F(x_2,p)\big) + \delta M_{i,1}^F(x_2,p) \ ,$$
$$x_2' = x_2 \odot \exp\big(\delta R_{i,2}^F(x_1',p)\big) + \delta M_{i,2}^F(x_1',p) \ .$$

Clearly, F_i is a C^1-diffeomorphism on $\mathbb{R}^{2d}$ with $\mathrm{J}_{F_i}(x,p) = \exp\big(\delta R_{i,1}^F(x_2,p) + \delta R_{i,2}^F(x_1',p)\big)$.

Then, Ψ defined by (25) is a C^1-diffeomorphism whose Jacobian can be recursively computed. Then, the kernel $P(x,p,w), \mathrm{d}(y,q,w))$ given by

$$\bar{\alpha}(x,p,v)\updelta_{\Psi^v(x,p)}(\mathrm{d}(y,q))\updelta_v(\mathrm{d}w) + (1 - \bar{\alpha}(x,p,v))\updelta_{(x,p,-v)}(\mathrm{d}(y,q,w))$$

where $\bar{\alpha}$ is defined in (24) is (π, S)-reversible. This kernel should be combined with (possibly partial) refreshment steps as discussed in Section 3.1; see Appendix D.6 for details.

Acknowledgments

AD and EM acknowledge support of the Lagrange Mathematical and Computing Research Center. Part of the paper was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project -5-100. It was also supported by the ANR ANR-19-CHIA-0002-01 "Chaire d'excellence en IA", projet SCAI.

Bibliography

Albergo, M., Kanwar, G., & Shanahan, P. (2019). Flow-based generative models for Markov chain Monte Carlo in lattice field theory. *Physical Review D*, 100(3), 034515.

Andrieu, C. & Livingstone, S. (2019). Peskun-Tierney ordering for Markov chain and process Monte Carlo: beyond the reversible scenario. *Accepted for publication in Ann. Statist.*

Ardizzone, L., Lüth, C., Kruse, J., Rother, C., & Köthe, U. (2019). Guided image generation with conditional invertible neural networks.

Baptista, R., Zahm, O., & Marzouk, Y. (2020). An adaptive transport framework for joint and conditional density estimation. *arXiv preprint arXiv:2009.10303*.

Bierkens, J. & Roberts, G. (2017). A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *Ann. Appl. Probab.*, 27(2), 846–882.

Chen, F., Lovász, L., & Pak, I. (1999). Lifting Markov chains to speed up mixing. In *Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999)* (pp. 275–281). ACM, New York.

Dalalyan, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3), 651–676.

Dalalyan, A. S. & Karagulyan, A. (2019). User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 129(12), 5278–5311.

Diaconis, P., Holmes, S., & Neal, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, 10(3), 726–752.

Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2017). Density estimation using Real NVP.

Douc, R., Moulines, E., Priouret, P., & Soulier, P. (2018). *Markov chains*. Springer Series in Operations Research and Financial Engineering. Springer, Cham.

Duane, S., Kennedy, A., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2), 216 – 222.

Durmus, A. & Moulines, E. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3), 1551–1587.

Durmus, A., Moulines, E., & Saksman, E. (2017). On the convergence of Hamiltonian Monte Carlo. *Accepted for publication in Ann. Statist.*

Fang, Y., Sanz-Serna, J.-M., & Skeel, R. D. (2014). Compressible generalized hybrid Monte Carlo. *The Journal of chemical physics*, 140(17), 174108.

Gustafson, P. (1998). A guided walk Metropolis algorithm. *Statistics and computing*, 8(4), 357–364.

Horowitz, A. M. (1991). A generalized guided Monte Carlo algorithm. *Physics Letters B*, 268(2), 247–252.

Hukushima, K. & Sakai, Y. (2013). An irreversible Markov-chain Monte Carlo method with skew detailed balance conditions. In *Journal of Physics: Conference Series*, volume 473 (pp. 012012).: IOP Publishing.

Levy, D., Hoffman, M. D., & Sohl-Dickstein, J. (2017). Generalizing Hamiltonian Monte Carlo with neural networks. *arXiv preprint arXiv:1711.09268*.

Ma, Y.-A., Chen, T., Wu, L., & Fox, E. B. (2016). A unifying framework for devising efficient and irreversible MCMC samplers. *arXiv preprint arXiv:1608.05973*.

Mengersen, K. & Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24, 101–121.

Michel, M. (2016). *Irreversible Markov chains by the factorized Metropolis filter : algorithms and applications in particle systems and spin models*. PhD thesis. Thèse de doctorat dirigée par Krauth, Werner Physique Paris Sciences et Lettres (ComUE) 2016.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, (pp. 113–162).

Neklyudov, K., Welling, M., Egorov, E., & Vetrov, D. (2020). Involutive MCMC: a unifying framework. *arXiv preprint arXiv:2006.16653*.

Ottobre, M. (2016). Markov chain Monte Carlo and irreversibility. *Reports on Mathematical Physics*, 77(3), 267–292.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2019). Normalizing flows for probabilistic modeling and inference. *arXiv preprint arXiv:1912.02762*.

Prangle, D. (2019). Distilling importance sampling. *arXiv preprint arXiv:1910.03632*.

Sherlock, C. & Thiery, A. H. (2019). A discrete bouncy particle sampler. *arXiv preprint 1707.05200*.

Sohl-Dickstein, J., Mudigonda, M., & DeWeese, M. R. (2014). Hamiltonian Monte Carlo without detailed balance. *arXiv preprint arXiv:1409.5191*.

Song, J., Zhao, S., & Ermon, S. (2017). A-NICE-MC: Adversarial training for MCMC. In *Advances in Neural Information Processing Systems* (pp. 5140–5150).

Spanbauer, S., Freer, C., & Mansinghka, V. (2020). Deep involutive generative models for neural mcmc. *arXiv preprint arXiv:2006.15167*.

Thin, A., Kotelevskii, N., Denain, J.-S., Grinsztajn, L., Durmus, A., Panov, M., & Moulines, E. (2020). Metflow: A new efficient method for bridging the gap between Markov chain Monte Carlo and variational inference. *arXiv preprint arXiv:2002.12253*.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4), 1701–1728.

Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1), 1–9.

Turitsyn, K. S., Chertkov, M., & Vucelja, M. (2011). Irreversible Monte Carlo algorithms for efficient sampling. *Physica D: Nonlinear Phenomena*, 240(4), 410 – 414.

A Notations, definitions and general Markov chain theory

In this section, we recall some basic facts and notations in a form that is useful for establishing properties of Markov chains. Let $(\mathsf{Z}, \mathcal{Z})$ be a measurable space where $\mathcal{Z}$ is a countably generated σ-algebra.

Definition 12 (Kernel). *A* kernel *on $\mathsf{Z} \times \mathcal{Z}$ is a map $P\colon \mathsf{Z} \times \mathcal{Z} \to \mathbb{R}_+$ such that*

(i) for any $A \in \mathcal{Z}$, $z \mapsto P(z, A)$ is measurable;

(ii) for any $z \in \mathsf{Z}$, the function $A \mapsto P(z, A)$ is a finite measure on $\mathcal{Z}$.

Definition 13 (Markov and sub-Markovian kernel). *A kernel P is Markovian (or P is a Markov kernel) if $P(z, \mathsf{Z}) = 1$ for all $z \in \mathsf{Z}$. A kernel P is submarkovian (or P is a sub-Markov kernel) if $P(z, \mathsf{Z}) \leqslant 1$ for all $z \in \mathsf{Z}$.*

For $f\colon \mathsf{Z} \to \mathbb{R}$ a measurable function, ν a probability distribution, and P a kernel on $\mathsf{Z} \times \mathcal{Z}$, we let $\nu(f) \overset{:=}{} \int f(z)\nu(\mathrm{d}z)$ and denote for $(z, A) \in \mathsf{Z} \times \mathcal{Z}$,

$$\nu P(A) = \int \nu(\mathrm{d}z) P(z, A) \;, \quad Pf(z) = \int P(z, \mathrm{d}z') f(z') \;.$$

Further, for $(z, A) \in \mathsf{Z} \times \mathcal{Z}$ define recursively for $n \geqslant 2$: $P^n(z, A) = \int P^{n-1}(z, \mathrm{d}z') P(z', A)$.

Definition 14 (Total variation distance). *For μ, ν two probability distributions on $(\mathsf{Z}, \mathcal{Z})$ we define the total variation distance between μ and ν by $\|\mu - \nu\|_{\mathrm{TV}} := \sup_{|f| \leqslant 1} |\mu(f) - \nu(f)|$, where the supremum is taken over the measurable function $f\colon \mathsf{Z} \to \mathbb{R}$.*

Definition 15 (Harmonic function). *Let P be a kernel on $(\mathsf{Z}, \mathcal{Z})$. Then a non-negative measurable function $h : \mathsf{Z} \to \mathbb{R}$ is said to be* harmonic *if $Ph = h$.*

Definition 16 (Irreducibility). *Let ν be a non trivial σ-finite measure on $\big(\mathsf{Z}, \mathcal{Z}\big)$. A kernel P is said to be ν-irreducible if for all $(z, A) \in \mathsf{Z} \times \mathcal{Z}$ such that $\nu(A) > 0$ there exists $n = n(z, A) \in \mathbb{N}$ such that $P^n(z, A) > 0$.*

Definition 17 (Periodicity and Aperiodicity). *P is periodic if there exists $n \in \mathbb{N}$, $n \geqslant 2$, and $A_i \in \mathcal{Z}$ for $i \in 1, \ldots, n$, non-empty and disjoint, such that for $z \in A_i$, $P(z, A_{i+1}) = 1$ with the convention $A_{n+1} = A_1$. Aperiodicity is the negation of periodicity.*

General Markov chain theory provides us with powerful tools to establish validity and convergence of MCMC algorithms, leading to basic convergence theorems such as those found in (Tierney, 1994, Theorem 1 and 3) and distilled below. We informally comment on the result below.

Theorem 18 (Tierney (1994)). *Suppose P is such that $\pi P = P$ and is $\pi-$irreducible. Then π is the unique invariant probability distribution of P and for any $f\colon \mathsf{Z} \to \mathbb{R}$ such that $\pi(|f|) < \infty$*

$$\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} f(Z_i) = \pi(f) \;, \tag{26}$$

almost surely for $\pi-$almost all $z \in \mathsf{Z}$. If in addition P is aperiodic then for $\pi-$almost all $z \in \mathsf{Z}$

$$\lim_{n \to \infty} \|P^n(z, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} = 0 \;. \tag{27}$$

The result is fairly intuitive. Invariance of π is a fixed point property ensuring that if $Z_i \sim \pi$ then $Z_{i+1} \sim \pi$. $\pi-$irreducibility simply says that the Markov chain should be able to reach any set of $\pi-$positive probability from any $z \in \mathsf{Z}$ in a finite number of iterations. Periodicity would clearly prevent (27) since the Markov chain would then periodically avoid visiting sets of positive $\pi-$probability. Averaging in (26) removes the need for this property. We note that establishing these properties is often overlooked and a necessary prerequisite to any more refined analysis characterising their performance, such as quantitative finite time convergence bounds as found for example in Dalalyan (2017); Dalalyan & Karagulyan (2019); Durmus & Moulines (2017).

B Standard reversible MH

We summarize in this Section the results presented in (Tierney, 1998, Section 2).

Definition 19 (Reversible kernel)**.** *A sub-Markovian kernel P on $(\mathsf{Z}, \mathcal{Z})$, P is π-reversible if and only if*

$$\check{\nu}\big(\mathrm{d}(z,z')\big) = \check{\nu}^F\big(\mathrm{d}(z,z')\big) ,$$

where $\check{\nu}\big(\mathrm{d}(z,z')\big) = \pi(\mathrm{d}z)P(z,\mathrm{d}z')$ and $\check{\nu}^F\big(\mathrm{d}(z,z')\big) = F_{\#}\nu\big(\mathrm{d}(z,z')\big) = \pi(\mathrm{d}z')P(z',\mathrm{d}z)$ is the push-foward measure of ν by $F\colon (z,z') \mapsto (z',z)$.

From a proposal Markov kernel Q, the MH method consists of considering a sub-Markovian kernel $Q_\alpha(z,\mathrm{d}z') = \alpha(z,z')Q(z,\mathrm{d}z')$. If π and Q admit a common dominating σ-finite measure μ on Z, such that $\pi(\mathrm{d}z) = \pi(z)\mu(\mathrm{d}z)$ (we use the same notation for the probability and the density) and $Q(z,\mathrm{d}z') = q(z,z')\mu(\mathrm{d}z')$, Q_α is π-reversible if

$$\alpha(z,z') = \begin{cases} \mathsf{a}\left(\frac{\pi(z')q(z',z)}{\pi(z)q(z,z')}\right) & \pi(z)q(z,z') > 0 \, , \\ 1 & \text{otherwise} \, , \end{cases}$$

where for any $t \in \mathbb{R}_+^*$,

$$t\mathsf{a}(1/t) = \mathsf{a}(t) \, .$$

We may take for example $\mathsf{a}(t) = \min(1,t)$ or $\mathsf{a}(t) = t/(1+t)$ which correspond to the classical Metropolis-Hastings and Barker ratio, respectively. To obtain a π-reversible Markov kernel P, it suffices to add a Dirac mass, *i.e.*

$$P(z,\mathrm{d}z') = Q_\alpha(z,\mathrm{d}z') + \big(1 - Q_\alpha(z,\mathsf{Z})\big)\delta_z(\mathrm{d}z') \, .$$

This construction can be generalized to the case where π or Q do not admit a density. In particular, let Φ be an invertible mapping on Z satisfying $\Phi^{-1} = \Phi$ (*i.e.* Φ is an involution) and consider $Q(z,\mathrm{d}z') = \delta_{\Phi(z)}(\mathrm{d}z')$ (when the current state is z, then the proposal is $\Phi(z)$). Define the measure $\nu = \pi + \Phi_{\#}\pi$ and denote by $h(z) = \mathrm{d}\pi/\mathrm{d}\nu(z)$ (h is the density of π w.r.t. ν). Then, $h\big(\Phi(z)\big)$ is a density of $\Phi_{\#}\pi$ w.r.t. ν. Denote $A = \{z \in \mathsf{Z} \, : \, h(z) \times h \circ \Phi(z) > 0\}$. Detailed balance holds if and only if for π-almost all $z \in A$ (see Tierney (1998)):

$$\alpha\big(z,\Phi(z)\big)h(z)/h \circ \Phi(z) = \alpha\big(\Phi(z), z\big) \, .$$

If $\mathsf{Z} = \mathbb{R}^d$ and ν is the Lebesgue measure, we obtain $\alpha\big(z,\Phi(z)\big) = \bar{\alpha}(z)$, where

$$\bar{\alpha}(z) = \mathsf{a}\left(\frac{\pi \circ \Phi(z)}{\pi(z)} \, \mathrm{J}_\Phi(z)\right) \, .$$

C Proofs of Section 2

C.1 Proof of (3)

Let $f\colon \mathsf{Z}^2 \to \mathbb{R}_+$ be a measurable function. The condition $\check{\mu}_P = \check{\mu}_P^s$ implies

$$I = \iint \check{\mu}_P\big(\mathrm{d}(z,z')\big) f(z,z') = \iint \check{\mu}_P\big(\mathrm{d}(z,z')\big) f\big(s(z'),s(z)\big) = \iint \pi(\mathrm{d}z) P(z,\mathrm{d}z') f\big(s(z'),s(z)\big) \; .$$

Using the change of variable $\tilde{z}' = s(z)$ and since s is an involution, we get

$$I = \iint s_\# \pi(\mathrm{d}\tilde{z}') P\big(s(\tilde{z}'),\mathrm{d}z'\big) f\big(s(z'),\tilde{z}'\big) \; .$$

Applying now the change of variable $\tilde{z} = s(z')$, we finally obtain

$$I = \iint s_\# \pi(\mathrm{d}\tilde{z}') s_\# P\big(s(\tilde{z}'),\mathrm{d}\tilde{z}\big) f(\tilde{z},\tilde{z}') \; .$$

Note that, for any $z \in \mathsf{Z}$ and $A \in \mathcal{Z}$,

$$s_\# P(z,A) = \int P(z,\mathrm{d}z') 1_A\big(s(z')\big) = PS(z,A) \; ,$$

showing that

$$I = \iint s_\# \pi(\mathrm{d}\tilde{z}') PS\big(s(\tilde{z}'),\mathrm{d}\tilde{z}\big) f(\tilde{z},\tilde{z}') = \iint s_\# \pi(\mathrm{d}\tilde{z}') SPS(\tilde{z}',\mathrm{d}\tilde{z}) f(\tilde{z},\tilde{z}'),$$

where we have used $\int SPg(\tilde{z}',\mathrm{d}\tilde{z}) = Pg\big(s(\tilde{z}')\big)$.

C.2 Proof of Proposition 3

We set $\check{\lambda} = \check{\nu} + \check{\nu}^s$. Note that $\check{\nu}$ and $\check{\nu}^s$ are absolutely continuous w.r.t. to $\check{\lambda}$. Denote by $\check{\lambda}^s = (F_s)_\# \check{\lambda}$ the pushforward of $\check{\lambda}$ by the transform $F_s(z,z') = \big(s(z'),s(z)\big)$: for any $C \in \mathcal{Z}^{\otimes 2}$

$$\check{\lambda}^s(C) = \int 1_C\big(s(z'),s(z)\big) \check{\lambda}\big(\mathrm{d}(z,z')\big) \; .$$

Since $(\check{\nu}^s)^s = \check{\nu}$, $\check{\lambda} = \check{\lambda}^s$. This implies, for any measurable function $f\colon \mathsf{Z}^2 \to \mathbb{R}_+$,

$$\iint f(z,z') \check{\lambda}\big(\mathrm{d}(z,z')\big) = \iint f\big(s(z'),s(z)\big) \check{\lambda}\big(\mathrm{d}(z,z')\big) \; .$$

We choose h to be a version of the Radon-Nikodym derivative $\mathrm{d}\check{\nu}/\mathrm{d}\check{\lambda}$ (the function is defined up to $\check{\lambda}$-negligible sets). Then by definition of $\check{\nu}^s$,

$$\begin{aligned}
\iint f(z,z') \check{\nu}^s\big(\mathrm{d}(z,z')\big) &= \iint f\big(s(z'),s(z)\big) \check{\nu}\big(\mathrm{d}(z,z')\big) = \iint f\big(s(z'),s(z)\big) h(z,z') \check{\lambda}\big(\mathrm{d}(z,z')\big) \\
&= \iint f\big(s(z'),s(z)\big) h(z,z') \check{\nu}\big(\mathrm{d}(z,z')\big) + \iint f\big(s(z'),s(z)\big) h(z,z') \check{\nu}^s\big(\mathrm{d}(z,z')\big) \\
&= \iint f(z,z') h\big(s(z'),s(z)\big) \check{\nu}^s\big(\mathrm{d}(z,z')\big) + \iint f(z,z') h\big(s(z'),s(z)\big) \check{\nu}\big(\mathrm{d}(z,z')\big) \\
&= \iint f(z,z') h\big(s(z'),s(z)\big) \check{\lambda}\big(\mathrm{d}(z,z')\big) \; ,
\end{aligned}$$

showing that

$$h\big(s(z'), s(z)\big) = \frac{\mathrm{d}\check{\nu}^s}{\mathrm{d}\check{\lambda}}(z, z') .$$

We then define

$$A_{\check{\nu}} = \big\{(z, z') \in \mathsf{Z}^2 \colon h(z, z') \times h\big(s(z'), s(z)\big) > 0\big\} .$$

In other words, if $(z, z') \notin A_{\check{\nu}}$, then either $h(z, z') = 0$ or $h\big(s(z'), s(z)\big) = 0$. Therefore, $\check{\nu}_{A,\mathrm{c}}$ and $\check{\nu}^s_{A,\mathrm{c}}$ are singular since $B_1 = \{(z, z') \in \mathsf{Z}^2 \colon h(z, z') > 0\}$, $B_2 = \{(z, z') \in \mathsf{Z}^2 \colon h\big(s(z'), s(z)\big) > 0\}$ are disjoint subsets of $A^{\mathrm{c}}_{\check{\nu}}$ and $\check{\nu}(B_2) = 0$, $\check{\nu}(B_1) = 0$. In addition, since for any set $B \in \mathsf{Z}^2$,

$$1_{A_{\check{\nu}} \cap B} h = 0 \ \check{\lambda} - \text{a.e. if and only if } 1_{A_{\check{\nu}} \cap B} h^s = 0 \ \check{\lambda} - \text{a.e.}$$

the restrictions $\check{\nu}_A$ and $\check{\nu}^s_A$ are equivalent. In addition,

$$\frac{\mathrm{d}\check{\nu}_A}{\mathrm{d}\check{\nu}^s}(z, z') = \frac{h(z, z')}{h\big(s(z'), s(z)\big)} = r(z, z') , \quad (z, z') \in A_{\check{\nu}} ,$$

satisfying $r(z, z') = 1/r\big(s(z'), s(z)\big)$.

C.3 Proof of Theorem 4

Define the σ-finite measure $\check{\rho}\big(\mathrm{d}(z, z')\big) = \alpha(z, z')\check{\nu}\big(\mathrm{d}(z, z')\big)$ and denote by $\check{\rho}^s = (F_s)_{\#}\check{\rho}$ the push-forward of $\check{\rho}$ by the transform $F_s(z, z') = \big(s(z'), s(z)\big)$: for any $C \in \mathcal{Z}^{\otimes 2}$

$$\check{\rho}^s(C) = \int 1_C\big(s(z'), s(z)\big)\check{\rho}\big(\mathrm{d}(z, z')\big) .$$

Note by definition of $\check{\nu}^s$ that

$$\check{\rho}^s\big(\mathrm{d}(z, z')\big) = \alpha\big(s(z'), s(z)\big)\check{\nu}^s\big(\mathrm{d}(z, z')\big) \tag{28}$$

We show below that under the stated assumptions $\check{\rho} = \check{\rho}^s$.

Define the function $\tilde{\alpha}(z, z') = \alpha\big(s(z'), s(z)\big)$. Since the set $A_{\check{\nu}}$ is s-symmetric, the set $A^{\mathrm{c}}_{\check{\nu}}$ is also s-symmetric and $\check{\nu}(\{(z, z') \in A^{\mathrm{c}}_{\check{\nu}} \, : \, \tilde{\alpha}(z, z') > 0\}) = 0$ using (i). Hence $\check{\rho}(A^{\mathrm{c}}_{\check{\nu}}) = \check{\rho}^s(A^{\mathrm{c}}_{\check{\nu}}) = 0$.

We have by Proposition 3 and (ii),

$$\begin{aligned} 1_{A_{\check{\nu}}}(z, z')\check{\rho}\big(\mathrm{d}(z, z')\big) &= 1_{A_{\check{\nu}}}(z, z')\alpha(z, z')\check{\nu}\big(\mathrm{d}(z, z')\big) = \alpha(z, z')r(z, z')\check{\nu}^s\big(\mathrm{d}(z, z')\big) \\ &= 1_{A_{\check{\nu}}}(z, z')\alpha\big(s(z'), s(z)\big)\nu^s\big(\mathrm{d}(z, z')\big) = 1_{A_{\check{\nu}}}(z, z')\check{\rho}^s\big(\mathrm{d}(z, z')\big) . \end{aligned} \tag{29}$$

Conversely, assume that $\check{\rho} = \check{\rho}^s$. Since by Proposition 3, $\check{\nu}_{A,\mathrm{c}}$ and $\check{\nu}^s_{A,\mathrm{c}}$ are mutually singular, there exist $B_1, B_2 \subset A^{\mathrm{c}}_{\check{\nu}}$ (see also the proof Proposition 3) such that $\check{\nu}_{A,\mathrm{c}}(B_2) = 0$ and $\check{\nu}^s_{A,\mathrm{c}}(B_1) = 0$. Therefore, we obtain using that $\check{\rho} = \check{\rho}^s$ and (28) that

$$\check{\rho}(A^{\mathrm{c}}_{\check{\nu}} \cap B_1) = \check{\rho}^s(A^{\mathrm{c}}_{\check{\nu}} \cap B_1) = 0 .$$

This result and $\check{\rho}(A^{\mathrm{c}}_{\check{\nu}} \cap B_2)$ imply $\check{\rho}(A^{\mathrm{c}}_{\check{\nu}}) = 0$ and therefore $\check{\nu}(\{(x, z') \in A^{\mathrm{c}}_{\check{\nu}} \, : \, \alpha(z, z') > 0\}) = 0$ showing (i). Finally, under the condition $\check{\rho} = \check{\rho}^s$ and Proposition 3, (29) holds and (ii) follows.

C.4 Checking the GMH rule (5)

We first check (i). By Proposition 3 and (5), $A_{\check{\nu}}^{c} = B_1 \cup B_2$ where $B_1 = \{(z, z') \in \mathsf{Z}^2 : h(z, z') = 0\}$ and $B_2 = \{(z, z') \in \mathsf{Z}^2 : h\big(s(z'), s(z)\big) = 0\}$, and for any $(z, z') \in B_2 \setminus B_1$, $\alpha(z, z') = 0$. Therefore, to show (i), it suffices to establish that $\check{\nu}(\{\alpha = 0\} \cap B_1) = 0$ which follows from

$$\check{\nu}(B_1) = \int 1_{B_1}(z, z')\check{\nu}\big(\mathrm{d}(z, z')\big) = \int 1_{B_1}(z, z')h(z, z')\check{\lambda}\big(\mathrm{d}(z, z')\big) = 0 .$$

We now check (ii). Note that by Proposition 3 and using that F_s is an involution, for $(z, z') \in A_{\check{\nu}}$,

$$\begin{aligned} \alpha(z, z')r(z, z') &= \mathsf{a}\big(1/r(z, z')\big)r(z, z') \\ &= \mathsf{a}\big(r(z, z')\big) = \alpha\big(s(z'), s(z)\big) . \end{aligned}$$

C.5 Expressions for a and b

We check the conditions on the nonnegative weights a and b so that the sub-Markovian kernel

$$R(z, \mathrm{d}z') = a(z)\delta_z(\mathrm{d}z') + b(z)\delta_{s(z)}(\mathrm{d}z')$$

is (π, S)-reversible. For f a nonnegative measurable function, we get

$$SRSf(z') = a\big(s(z')\big)f(z') + b\big(s(z')\big)f\big(s(z')\big) .$$

Hence, we obtain, for any nonnegative measurable function g,

$$\begin{aligned} \iint \pi(\mathrm{d}z')SRS(z', \mathrm{d}z)f(z)g(z') &= \int \pi(\mathrm{d}z')a\big(s(z')\big)f(z')g(z') + \int \pi(\mathrm{d}z')b\big(s(z')\big)f\big(s(z')\big)g(z') \\ &= \int \pi(\mathrm{d}z)a\big(s(z')\big)\delta_z(\mathrm{d}z')f(z)g(z') + \int \pi(\mathrm{d}z')b(z')f(z')g\big(s(z')\big) \\ &= \int \pi(\mathrm{d}z)a\big(s(z')\big)\delta_z(\mathrm{d}z')f(z)g(z') + \int \pi(\mathrm{d}z)b(z)f(z)\delta_{s(z)}(\mathrm{d}z')g(z') , \end{aligned}$$

where we have used $s_{\#}\pi = \pi$. The result implies that

$$\pi(\mathrm{d}z')SRS(z', \mathrm{d}z) = \pi(\mathrm{d}z)a\big(s(z)\big)\delta_z(\mathrm{d}z') + \pi(\mathrm{d}z)b(z)\delta_{s(z)}(\mathrm{d}z').$$

Therefore, (3) is satisfied (e.g. $\pi(\mathrm{d}z')SRS(z', \mathrm{d}z) = \pi(\mathrm{d}z)R(z, \mathrm{d}z')$) and R is (π, S)-reversible if $a(z) = a\big(s(z)\big)$.

In addition, the total mass of $Q_\alpha(z, \mathrm{d}z')$ is $Q_\alpha(z, \mathsf{Z})$. The missing mass is therefore $1 - Q_\alpha(z, \mathsf{Z})$. Since the total mass of R is $a(z) + b(z)$ we must have $a(z) + b(z) = 1 - Q_\alpha(z, \mathsf{Z})$.

We may for example set $a(z) = 0$ and $b(z) = 1 - Q_\alpha(z, \mathsf{Z})$, which coincides with the classical MH rule when $s = \mathrm{Id}$. We may also take $a(z) = 1 - Q_\alpha(z, \mathsf{Z}) - b(z)$ where b satisfies $0 \leqslant b(z) \leqslant 1 - Q_\alpha(z, \mathsf{Z})$ and $b(z) - b\big(s(z)\big) = Q_\alpha(s(z), \mathsf{Z}) - Q_\alpha(z, \mathsf{Z})$. As suggested in Turitsyn et al. (2011), we may set $b(z) = \max\big(0, Q_\alpha\big(s(z), \mathsf{Z}\big) - Q_\alpha(z, \mathsf{Z})\big)$ which is shown to be optimal w.r.t. to the Peskun ordering in Andrieu & Livingstone (2019). Note however that this choice for b is not always easily computable.

C.6 Applications of (5): case with densities

Note that $\check{\nu}\big(\mathrm{d}(z,z')\big) = \pi(z)q(z,z')\mu^{\otimes 2}\big(\mathrm{d}(z,z')\big)$, $\check{\nu}^s\big(\mathrm{d}(z,z')\big) = \pi\big(s(z')\big)q\big(s(z'),s(z)\big)\mu^{\otimes 2}\big(\mathrm{d}(z,z')\big)$, since $s_{\#}\mu = \mu$. In addition, $h = \tilde{h}/\{\tilde{h} + \tilde{h} \circ F_s\}$, $\tilde{h}(z,z') = \pi(z)q(z,z')$ and therefore $A_{\check{\nu}}$ in (8) is given by

$$A_{\check{\nu}} = \big\{\pi(z)q(z,z') \times \pi\big(s(z')\big)q\big(s(z'),s(z)\big) > 0\big\}\ ,$$

and for $(z,z') \in A_{\check{\nu}}$,

$$r(z,z') = \frac{\pi(z)q(z,z')}{\pi\big(s(z')\big)q\big(s(z'),s(z)\big)}\ .$$

Therefore, we obtain using (5) that

$$\alpha(z,z') = \begin{cases} \mathsf{a}\left[\frac{\pi\left(s(z')\right)q\left(s(z'),s(z)\right)}{\pi(z)q(z,z')}\right] & \pi(z)q(z,z') \neq 0, \\ 1 & \pi(z)q(z,z') = 0\ . \end{cases}$$

In addition, note that using $s_{\#}\pi = \pi$, $s_{\#}\mu = \mu$ and s is an involution, we obtain that

$$\pi = \pi \circ s\ .$$

Therefore, we obtain

$$\alpha(z,z') = \begin{cases} \mathsf{a}\left[\frac{\pi(z')q\left(s(z'),s(z)\right)}{\pi(z)q(z,z')}\right] & \pi(z)q(z,z') \neq 0, \\ 1 & \pi(z)q(z,z') = 0\ . \end{cases} \tag{30}$$

C.7 Proof of Theorem 6

We preface the proof by the following result. Define $\mathsf{Z}^+ = \{z \in \mathsf{Z} \colon \pi(z) > 0\}$ and set

$$P(z,\mathrm{d}z') = Q_\alpha(z,\mathrm{d}z') + \{1 - Q_\alpha(z,\mathsf{Z})\}\updelta_{s(z)}(\mathrm{d}z')\ . \tag{31}$$

where $Q_\alpha(z,z') = \alpha(z,z')Q(z,\mathrm{d}z')$ and α is given by (5). Note that P corresponds to (7) with $a \equiv 0$ and $b(z) = 1 - Q_\alpha(z,\mathsf{Z})$.

Proposition 20. *Consider P defined by (31). Assume that P is π-irreducible and $Q(z,\mathsf{Z}^+) = 1$ for any $z \notin \mathsf{Z}^+$. Further, suppose that π is not a Dirac mass. Then, P is Harris recurrent.*

Proof. Since π is invariant for P by Theorem 4, P is recurrent by (Douc et al., 2018, Theorem 10.1.6). Therefore, (Douc et al., 2018, Corollary 9.2.16, Proposition 5.2.12) show that for any bounded harmonic function $h\colon \mathsf{Z} \to \mathbb{R}$, *i.e.* satisfying $Ph = h$, $h = \pi(h)$, π a.e.. Then, if $A_h = \{h \neq \pi(h)\}$, $\pi(A_h) = 0$. By (Douc et al., 2018, Theorem 10.2.11), P is Harris recurrent if

$$h(z) = \pi(h) \text{ for any } z \in \mathsf{Z}\ . \tag{32}$$

First, consider $z \in \mathsf{Z}^+$. Define $B = \{z' \colon \pi(z')q\big(s(z'), s(z)\big) > 0\}$ and $C = \{z' \colon q(z,z') = 0\}$. Let A be a π-negligible set, $\pi(A) = 0$. Using $\mathsf{a}(t) \leqslant t$ by (6) for any $t \in \mathbb{R}_+^*$, $\pi(z) \neq 0$ and (30), we get

$$\begin{aligned}\int 1_A(z')\alpha(z,z')q(z,z')\mu(\mathrm{d}z') &= \int 1_{A\cap B\cap C}(z')\alpha(z,z')q(z,z')\mu(\mathrm{d}z') \\ &\quad + \int 1_{A\cap B\cap C^{\mathrm{c}}}(z')\mathsf{a}\left[\frac{\pi(z')q\big(s(z'),s(z)\big)}{\pi(z)q(z,z')}\right]q(z,z')\mu(\mathrm{d}z') \\ &\leqslant \int 1_{A\cap B\cap C^{\mathrm{c}}}(z')\frac{\pi(z')q\big(s(z'),s(z)\big)}{\pi(z)}\mu(\mathrm{d}z') \\ &\leqslant \int 1_{A\cap B\cap C^{\mathrm{c}}}(z')\frac{q\big(s(z'),s(z)\big)}{\pi(z)}\pi(\mathrm{d}z') = 0\;,\end{aligned} \tag{33}$$

where the last identity follows from $\pi(A) = 0$. Applying this identity with A_h yields to

$$\int \alpha(z,z')q(z,z')h(z')\mathrm{d}\mu(z') = \int 1_{A_h^{\mathrm{c}}}(z')\alpha(z,z')q(z,z')h(z')\mathrm{d}\mu(z') = \pi(h)Q_\alpha(z,\mathsf{Z})\;.$$

Therefore, the condition $Ph(z) = h(z)$ for any $z \in \mathsf{Z}$ and (31) imply that

$$h(z) = \pi(h)Q_\alpha(z,\mathsf{Z}) + h\circ s(z)\{1 - Q_\alpha(z,\mathsf{Z})\}\;.$$

Applying P to the previous equation, we obtain, denoting $\bar{\alpha}(z) = Q_\alpha(z,\mathsf{Z})$

$$\begin{aligned}h(z) = Ph(z) &= \pi(h)\{Q_\alpha\bar{\alpha}(z) + \{1-\bar{\alpha}(z)\}\bar{\alpha}\circ s(z)\} \\ &\quad + \int Q_\alpha(z,\mathrm{d}z')h\circ s(z')\{1-\bar{\alpha}(z')\} + h(z)\{1-\bar{\alpha}(z)\}\big\{1-\bar{\alpha}\big(s(z)\big)\big\}\;.\end{aligned} \tag{34}$$

Denote $A_{h\circ s} = \{z \in \mathsf{Z} \,:\, h\circ s(z) \neq \pi(h)\}$. Note that, $\pi(A_{h\circ s}) = s_\#\pi(A_h) = \pi(A_h) = 0$. Using (33), we get for $z \in \mathsf{Z}^+$, $Q_\alpha(z, A_{h\circ s}) = 0$, which implies

$$\begin{aligned}\int Q_\alpha(z,\mathrm{d}z')h\circ s(z')\{1-\bar{\alpha}(z')\} &= \int 1_{A_{h\circ s}^{\mathrm{c}}}(z')Q_\alpha(z,\mathrm{d}z')h\circ s(z')\{1-\bar{\alpha}(z')\} \\ &= \pi(h)\int Q_\alpha(z,\mathrm{d}z')\{1-\bar{\alpha}(z')\}\;.\end{aligned}$$

Plugging this relation into (34) we obtain

$$h(z) = \pi(h)[Q_\alpha\bar{\alpha}(z) + \{1-\bar{\alpha}(z)\}\bar{\alpha}\circ s(z)] + \pi(h)[\bar{\alpha}(z) - Q_\alpha\bar{\alpha}(z)] + h(z)\{1-\bar{\alpha}(z)\}\{1-\bar{\alpha}\circ s(z)\}\;.$$

Using straightforward algebra, the previous identity implies

$$\{\pi(h) - h(z)\}\{\bar{\alpha}(z) + \bar{\alpha}\circ s(z) - \bar{\alpha}(z)\times\bar{\alpha}\circ s(z)\} = 0.$$

Since P is π-irreducible and π is not a Dirac mass, $\bar{\alpha}(z) \neq 0$, we get that for all $z \in \mathsf{Z}^+$,

$$h(z) = \pi(h)\;. \tag{35}$$

Consider now the case $z \notin \mathsf{Z}_+$. Using that $Q(z, \mathsf{Z}_+)$ by assumption and $\alpha(z, z') = 1$ by (30) for any $z' \in \mathsf{Z}$, we get

$$h(z) = Ph(z) = \int_{\mathsf{Z}_+} q(z, z')h(z')\mu(z') = \int_{\mathsf{Z}_+} \{q(z, z')h(z')/\pi(z')\}\pi(z')\mu(z') = \pi(h) .$$

Combining this result with (35) completes the proof of (32). $\square$

Proposition 21. *Assume the conditions of Theorem 6. Then for any $A \in \mathcal{Z}$ such that $\pi(A) > 0$, we have*

$$P(z, A) > 0 \quad \textit{for any } z \in \mathsf{Z} . \tag{36}$$

Proof. Consider first the case $z \in \mathsf{Z}_+ = \{z \in \mathsf{Z} : \pi(z) > 0\}$. Then, by (30) and the condition if $\pi(z') > 0$, then $q(\tilde{z}, z') \times q\big(s(\tilde{z}), s(z')\big) > 0$ for any $\tilde{z} \in \mathsf{Z}$, we have

$$P(z, A) \geqslant \int 1_{A \cap \mathsf{Z}_+}(z')\mathsf{a}\left[\frac{\pi(z')q\big(s(z'), s(z)\big)}{\pi(z)q(z, z')}\right] q(z, z')\mu(\mathrm{d}z') > 0 ,$$

since $\pi(A) > 0$ implies that $\mu(A \cap \mathsf{Z}_+) > 0$. Second consider the case $z \notin \mathsf{Z}_+$. Then, $\alpha(z, z') = 1$ for any $z' \in \mathsf{Z}$ and we get

$$P(z, A) \geqslant \int 1_{A \cap \mathsf{Z}_+}(z')q(z, z')\mu(\mathrm{d}z') > 0 ,$$

which concludes the proof of (36). $\square$

Proof of Theorem 6. π-irreducibility of P follows from Proposition 21. We show that P is π-irreducible and aperiodic. Indeed, this result and (Douc et al., 2018, Theorem 7.2.1, Theorem 11.3.1) imply (27) for all $z \in \mathsf{Z}$. Finally, (Douc et al., 2018, Corollary 9.2.16, Proposition 5.2.14) establish (26) for all $z \in \mathsf{Z}$.

The fact that P is aperiodic is a direct consequence of (36) and (Douc et al., 2018, Theorem 9.3.10). $\square$

C.8 Proofs of (11) and (12)

Consider $\check{\nu}\big(\mathrm{d}(z, z')\big) = \pi(\mathrm{d}z)\delta_{\Phi(z)}(\mathrm{d}z')$, where Φ is an invertible mapping on Z satisfying $\Phi^{-1} = s \circ \Phi \circ s$. For any measurable function $f : \mathsf{Z}^2 \to \mathbb{R}_+$, we get

$$\begin{aligned} \iint \check{\nu}^s\big(\mathrm{d}(z, z')\big)f(z, z') &= \iint \check{\nu}\big(\mathrm{d}(z, z')\big)f\big(s(z'), s(z)\big) = \int \pi(\mathrm{d}z)f\big(s \circ \Phi(z), s(z)\big) \\ &= \int s_\# \pi(\mathrm{d}z')f(s \circ \Phi \circ s(z'), z') = \iint \pi(\mathrm{d}z')\delta_{\Phi^{-1}(z')}(\mathrm{d}z)f(z, z') . \end{aligned}$$

Define

$$\check{\lambda}\big(\mathrm{d}(z, z')\big) = \check{\nu}\big(\mathrm{d}(z, z')\big) + \check{\nu}^s\big(\mathrm{d}(z, z')\big) = \pi(\mathrm{d}z)\delta_{\Phi(z)}(\mathrm{d}z') + \pi(\mathrm{d}z')\delta_{\Phi^{-1}(z')}(\mathrm{d}z) .$$

Set $\lambda = \pi + \Phi^{-1}_{\#}\pi$ and define $k(z) = (\mathrm{d}\pi/\mathrm{d}\lambda)(z)$. Note that for any measurable function $f \colon \mathsf{Z}^2 \to \mathbb{R}_+$,

$$
\begin{aligned}
\int f(z,z')\check{\lambda}\big(\mathrm{d}(z,z')\big) &= \int \pi(\mathrm{d}z) f\big(z,\Phi(z)\big) + \int \pi(\mathrm{d}z') f\big(\Phi^{-1}(z'), \Phi \circ \Phi^{-1}(z')\big) \\
&= \int f\big(z,\Phi(z)\big)\lambda(\mathrm{d}z) \,.
\end{aligned}
$$

Then, for any measurable function $f \colon \mathsf{Z}^2 \to \mathbb{R}_+$, we get since $k(z) = \mathrm{d}\pi/\mathrm{d}\lambda(z)$,

$$
\iint \check{\nu}\big(\mathrm{d}(z,z')\big) f(z,z') = \int \pi(\mathrm{d}z) f\big(z,\Phi(z)\big) = \int k(z) f\big(z,\Phi(z)\big)\lambda(\mathrm{d}z) \,.
$$

On the other hand,

$$
\iint \check{\nu}\big(\mathrm{d}(z,z')\big) f(z,z') = \int h(z,z') f(z,z')\check{\lambda}\big(\mathrm{d}(z,z')\big) = \int h\big(z,\Phi(z)\big) f\big(z,\Phi(z)\big)\lambda(\mathrm{d}z) \,.
$$

showing that $h\big(z,\Phi(z)\big) = (\mathrm{d}\check{\nu}/\mathrm{d}\check{\lambda})\big(z,\Phi(z)\big) = k(z)$ λ-a.e. . Setting $z' = s(z)$ and using $\Phi^{-1} = s \circ \Phi \circ s$, we get

$$
h\big(s \circ \Phi(z), s(z)\big) = h(s \circ \Phi \circ s(z'), z') = h(\Phi^{-1}(z'), z')
$$

Setting now $z'' = \Phi^{-1}(z')$, *i.e.* $z'' = \Phi^{-1} \circ s(z) = s \circ \Phi(z)$, we finally obtain

$$
h\big(s \circ \Phi(z), s(z)\big) = h\big(z'', \Phi(z'')\big) = k(z'') = k\big(\Phi^{-1} \circ s(z)\big) = k\big(s \circ \Phi(z)\big) \,.
$$

The proof of (11) and (12) is concluded using Theorem 4.

C.9 Proof of (13)

We now consider the case $\mathsf{Z} = \mathbb{R}^d$ and $\pi(\mathrm{d}z) = \pi(z)\mathrm{d}z$. We first identify the dominating measure λ defined in (11). For any nonnegative measurable function f,

$$
\begin{aligned}
\lambda(f) &= \int f(z)\pi(z)\mathrm{d}z + \int f \circ \Phi^{-1}(z)\pi(z)\mathrm{d}z \\
&= \int f(z)\pi(z)\mathrm{d}z + \int f(z)\pi \circ \Phi(z)\,\mathrm{J}_{\Phi}(z)\mathrm{d}z \,.
\end{aligned}
$$

Hence, $\lambda(\mathrm{d}z) = \lambda(z)\mathrm{d}z$ with

$$
\lambda(z) = \pi(z) + \pi \circ \Phi(z)\,\mathrm{J}_{\Phi}(z) \,.
$$

Plugging this expression in (11), we get that

$$
k(z) = \frac{\mathrm{d}\pi}{\mathrm{d}\lambda}(z) = \frac{\pi(z)}{\pi(z) + \pi \circ \Phi(z)\,\mathrm{J}_{\Phi}(z)} \,. \tag{37}
$$

We have for any function nonnegative measurable function f,

$$
s_{\#}\pi(f) = \int \pi(z) f \circ s(z)\mathrm{d}z = \int \pi \circ s(z)\,\mathrm{J}_{s}(z) f(z)\mathrm{d}z,
$$

which implies since $s_{\#}\pi = \pi$, that $\pi \circ s(z) = \pi(z)/\, \mathrm{J}_s(z)$ Leb_d-a.e.. Hence, we get that

$$
\begin{aligned}
k\big(s \circ \Phi(z)\big) &= \frac{\pi\big(s \circ \Phi(z)\big)}{\pi\big(s \circ \Phi(z)\big) + \pi\big(\Phi \circ s \circ \Phi(z)\big)\, \mathrm{J}_\Phi\big(s \circ \Phi(z)\big)} \\
&= \frac{\pi \circ \Phi(z)}{\pi \circ \Phi(z) + \pi(z)\rho_\Phi(z)}, \qquad (38)
\end{aligned}
$$

where we have set

$$
\rho_\Phi(z) = \frac{\mathrm{J}_\Phi\big(s \circ \Phi(z)\big)\, \mathrm{J}_s\big(\Phi(z)\big)}{\mathrm{J}_s(z)} \; .
$$

Since $\Phi \circ s \circ \Phi(z) = s(z)$, we get that

$$
\mathrm{J}_\Phi\big(s \circ \Phi(z)\big)\, \mathrm{J}_s\big(\Phi(z)\big)\, \mathrm{J}_\Phi(z) = \mathrm{J}_s(z),
$$

which implies that

$$
\rho_\Phi(z) = 1/\, \mathrm{J}_\Phi(z) \; .
$$

Plugging this expression into (38), we finally get that

$$
k\big(s \circ \Phi(z)\big) = \frac{\pi \circ \Phi(z)\, \mathrm{J}_\Phi(z)}{\pi(z) + \pi \circ \Phi(z)\, \mathrm{J}_\Phi(z)} \; .
$$

Combining this result with (12) and (37) concludes the proof of (13).

D Proofs of Section 3

D.1 Generalized Hamiltonian Monte Carlo algorithms

Consider the two following assumptions:

NICE1. *For any $i \in \{1, \ldots, m\}$, $N_{m+1-i} = M_i$.*

NICE2. *For any $i \in \{1, \ldots, m\}$, M_i and N_i are* L*-Lipschitz and $h \leqslant c_0/[\mathrm{L}^{1/2} m]$, where $c_0 \approx 0.3$.*

Lemma 22. *Assume **NICE**1. Then, $s \circ \Phi \circ s = \Phi^{-1}$.*[1]

Proof. Denote for $i \in \{1, \ldots, m\}$, $\mathrm{F}_i = \Psi_{N_i} \circ \Upsilon \circ \Psi_{M_i}$, where $\Upsilon(x, p) = (x + hp, p)$, $\Psi_M(x, p) = \big(x, p + hM(x)\big)$ and $\Psi_N(x, p) = \big(x, p + hN(x)\big)$. Each of those transforms verify

$$
s \circ \Upsilon \circ s = \Upsilon^{-1} \; , \; s \circ \Psi_M \circ s = \Psi_M^{-1} \; , \; s \circ \Psi_N \circ s = \Psi_N^{-1} \; .
$$

Then, $s \circ \mathrm{F}_i \circ s = \Psi_{N_i}^{-1} \circ \Upsilon^{-1} \circ \Psi_{M_i}^{-1}$ and thus,

$$
s \circ \Phi \circ s = \Psi_{N_m}^{-1} \circ \Upsilon^{-1} \circ \Psi_{M_m}^{-1} \circ \cdots \circ \Psi_{N_1}^{-1} \circ \Upsilon^{-1} \circ \Psi_{M_1}^{-1} \; .
$$

On the other hand,

$$
\Phi^{-1} = \mathrm{F}_1^{-1} \circ \cdots \circ \mathrm{F}_m^{-1} = \Psi_{M_1}^{-1} \circ \Upsilon^{-1} \circ \Psi_{N_1}^{-1} \circ \cdots \circ \Psi_{M_m}^{-1} \circ \Upsilon^{-1} \circ \Psi_{N_m}^{-1} \; .
$$

Applying **NICE**1 concludes the proof. ∎

[1] This condition is missing in the main text due to a late error with our versioning system.

D.1.1 Reversibility vs. persistency

[2] In Subsection 3.1 we define the deterministic Markov kernel on Z,

$$P\big((x,p);\mathrm{d}(x',p')\big)=\bar{\alpha}(x,p)\updelta_{\Phi(x,p)}\big(\mathrm{d}(x',p')\big)+\big(1-\bar{\alpha}(x,p)\big)\updelta_{(x,-p)}\big(\mathrm{d}(x',p')\big)\ ,$$

where $\bar{\alpha}(x,p)$ is given by (16). Such kernels are most likely not ergodic and the momentum must be refreshed in order to lead to an ergodic Markov chain $\big(Z_i=(X_i,P_i)\big)_{i\in\mathbb{N}}$. We focus on "full refreshment", that is the scenario where the momentum is drawn afresh from its stationary distribution before applying P, in which case it can be checked that $(X_i)_{i\in\mathbb{N}}$ is a Markov chain of (marginal) Markov kernel,

$$K(x,\mathrm{d}y)=K_\alpha(x,\mathrm{d}y)+\{1-K_\alpha(x,\mathbb{R}^d)\}\updelta_x(\mathrm{d}y)\ , \tag{39}$$

where $K_\alpha(x,\mathrm{d}y)=\int\bar{\alpha}(x,p)\varphi(p)\updelta_{G_x(p)}(\mathrm{d}y)\mathrm{d}p$ with $G_x(p)=\mathrm{proj}_1\circ\Phi(x,p)$, $\mathrm{proj}_1(x,p)=x$. Sampling from K is described in Algorithm 1. It is also the case that $(X_i)_{i\in\mathbb{N}}$ is time-reversible (see e.g. Duane et al. (1987)), which has the disadvantage of loosing the potentially advantageous persistency features of P. It is possible to recover persistency by considering the mixture of kernels

$$T:=\omega P+(1-\omega)L \tag{40}$$

for $\omega\in[0,1]$, where $L\big((x,p),\mathrm{d}(y,q)\big)=K(x,\mathrm{d}y)\varphi(q)\mathrm{d}q$. The kernel L refreshes independently the position x and the momentum p. Since the target distribution is the product of π_0 and φ (thus the position and the momentum are independent), it is easily checked as well that L leaves π-invariant since

$$\pi L\big(\mathrm{d}(y,q)\big)=\int\pi_0(\mathrm{d}x)K(x,\mathrm{d}y)\varphi(p)\mathrm{d}p\varphi(q)\mathrm{d}q=\pi_0(\mathrm{d}y)\varphi(q)\mathrm{d}q=\pi\big(\mathrm{d}(y,q)\big)\ .$$

Note further that L is π-reversible as K is π_0-reversible. In what follows we establish π_0-irreducibility of K, and in particular that Proposition 21 holds, which immediately implies $\pi_0\otimes(\varphi\times\mathrm{Leb})-$irreducibility of T and allows us to apply Theorem 18 and conclude about convergence. Sampling from T is described in Algorithm 2. In future work we will consider the scenario where p is updated using partial refreshment such as suggested in Horowitz (1991), for example by using an AR(1) process when φ is a normal distribution, which requires an extension of our results; see Algorithm 3.

D.1.2 Proof of (18) and Theorem 7

We first establish the elementary equation (18).

Lemma 23. *Assume that for each $x\in\mathbb{R}^d$, $G_x:p\mapsto\mathrm{proj}_1\circ\Phi(x,p)$ is a C^1-diffeomorphism. Then, $K_\alpha(x,\mathrm{d}y)$ has a density $K_\alpha(x,\mathrm{d}y)=\alpha(x,y)q(x,y)\mathrm{d}y$ where*

$$\alpha(x,y)=\mathsf{a}\left(\frac{\pi_0(y)\varphi\big\{H_x\big(G_x^{-1}(y)\big)\big\}}{\pi_0(x)\varphi\big(G_x^{-1}(y)\big)}\right)\ ,$$
$$q(x,y)=\varphi(G_x^{-1}(y))\,\mathrm{J}_{G_x^{-1}}(y)\ .$$

[2]We follow the order of the main text here, but this discussion should be after the proofs of the following subsections.

Proof. First by (16), for any $(x,p) \in \mathbb{R}^{2d}$, we have by definition

$$K_\alpha f(x) = \int \bar{\alpha}(x,p)\varphi(p) f\big(G_x(p)\big)\mathrm{d}p = \int \mathsf{a}\left(\frac{\pi \circ \Phi(x,p)}{\pi(x,p)}\right)\varphi(p) f\big(G_x(p)\big)\mathrm{d}p \,.$$

Then, using the change of variable $y = G_x(p)$, we obtain

$$K_\alpha f(x) = \int \mathsf{a}\left(\frac{\pi \circ \Phi\big(x, G_x^{-1}(y)\big)}{\pi\big(x, G_x^{-1}(y)\big)}\right)\varphi\big(G_x^{-1}(y)\big)\,\mathrm{J}_{G_x^{-1}}(y) f(y)\mathrm{d}y \,,$$

which concludes the proof of (18) since $\pi = \pi_0 \otimes \varphi$. □

We now prove Theorem 7 which gives conditions on the mappings $\{M_i, N_i\}_{i=1}^m$ that ensure that for all $x \in \mathbb{R}^d$, G_x is a C^1-diffeomorphism.

Theorem 24. *Assume* ***NICE2****. Then, for any $x \in \mathbb{R}^d$, the function $G_x(p) = \mathrm{proj}_1 \circ \Phi(x,p)$ is a C^1 diffeomorphism. Moreover, the GMH kernel based on NICE transitions is ergodic.*

We preface the proof by some auxiliary results. Recall that one step of the NICE transition is given by $\mathrm{F}_i(x_i, p_i) = (x_{i+1}, p_{i+1})$, where:

$$\begin{cases} p_{i+1/2} &= p_i + hM_i(x_i), \\ x_{i+1} &= x_i + hp_{i+1/2}, \\ p_{i+1} &= p_{i+1/2} + hN_i(x_{i+1}). \end{cases}$$

Denote

$$\Lambda^{(j)} = \mathrm{F}_j \circ \cdots \circ \mathrm{F}_1 \ .$$

Lemma 25. *For all $k \in \mathbb{N}^*$, we get*

$$\begin{aligned} x_k &= x_1 + (k-1)hp_1 + h^2\sum_{i=1}^{k-1}(k-i)M_i(x_i) + h^2\sum_{i=1}^{k-2}(k-1-i)N_i(x_{i+1}) \,, \\ p_k &= p_1 + h\sum_{i=1}^{k-1}M_i(x_i) + h\sum_{i=1}^{k-1}N_i(x_{i+1}) \,. \end{aligned}$$

Proof. The proof proceeds by induction. The assertion is obviously true for $k = 2$. Let us suppose that the assertion holds true for some $k \in \mathbb{N}^*$.

$$
\begin{aligned}
p_{k+1/2} &= p_k + hM_k(x_k) = p_1 + h\sum_{i=1}^{k} M_i(x_i) + h\sum_{i=1}^{k-1} N_i(x_{i+1}), \\
x_{k+1} &= x_k + hp_{k+1/2} \\
&= x_1 + (k-1)hp_1 + h^2\sum_{i=1}^{k-1}(k-i)M_i(x_i) + h^2\sum_{i=1}^{k-2}(k-1-i)N_i(x_{i+1}) \\
&\qquad + h\left(p_1 + h\sum_{i=1}^{k} M_i(x_i) + h\sum_{i=1}^{k-1} N_i(x_{i+1})\right) \\
&= x_1 + khp_1 + h^2\sum_{i=1}^{k}(k+1-i)M_i(x_i) + h^2\sum_{i=1}^{k-1}(k-i)N_i(x_{i+1}), \\
p_{k+1} &= p_{k+1/2} + hN_k(x_{k+1}) = p_1 + h\sum_{i=1}^{k} M_i(x_i) + h\sum_{i=1}^{k} N_i(x_{i+1}).
\end{aligned}
$$

This concludes the proof. □

Denote for all $(x_1, p_1) \in \mathbb{R}^{2d}$,

$$
\begin{aligned}
G_{x_1}(p_1) &= x_1 + mhp_1 + h^2\Theta_m(x_1, p_1), \\
\Theta_m(x_1, p_1) &= \sum_{i=1}^{m}(m+1-i)M_i(x_i) + \sum_{i=1}^{m-1}(m-i)N_i(x_{i+1}).
\end{aligned}
$$

Since the mappings $\{M_k\}_{k=1}^m$, $\{N_k\}_{k=1}^m$ are continuously differentiable, the mapping Ξ_k is continuously differentiable.

Lemma 26. *For $(x_1, p_1), (\tilde{x}_1, \tilde{p}_1) \in \mathbb{R}^{2d}$, denote $(x_{k+1}, p_{k+1}), (\tilde{x}_{k+1}, \tilde{p}_{k+1})$ the states obtained after k NICE-based transitions. Under the Lipschitz constraint L, we have*

$$
\|x_{k+1} - \tilde{x}_{k+1}\| + \mathrm{L}^{-1/2}\|p_{k+1} - \tilde{p}_{k+1}\| \leqslant \left\{1 + h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}\right)\right\}^k \left\{\|x_1 - \tilde{x}_1\| + \mathrm{L}^{-1/2}\|p_1 - \tilde{p}_1\|\right\} ,
$$

where $\vartheta_1(s) = 2 + s + s^2$.

Proof. We show this result for $k = 1$ and the apply a straightforward induction. For $k = 1$, we have

$$
\begin{aligned}
\|x_2 - \tilde{x}_2\| &= \left\|x_1 + h^2M_1(x_1) + hp_1 - \left\{\tilde{x}_1 + h^2M_1(\tilde{x}_1) + h\tilde{p}_1\right\}\right\| \\
&\leqslant \left(1 + h^2\mathrm{L}\right)\|x_1 - \tilde{x}_1\| + h\|p_1 - \tilde{p}_1\| \ .
\end{aligned}
$$

Moreover, we have

$$
\begin{aligned}
\|p_2 - \tilde{p}_2\| &= \|p_1 - \tilde{p}_1 - h\{N_1(x_2) + M_1(x_1)\} + h\{N_1(\tilde{x}_2) + M_1(\tilde{x}_1)\}\| \\
&\leqslant \|p_1 - \tilde{p}_1\| + h\mathrm{L}\{\|\tilde{x}_2 - x_2\| + \|\tilde{x}_1 - x_1\|\} \\
&\leqslant \left(1 + h^2\mathrm{L}\right)\|p_1 - \tilde{p}_1\| + h\mathrm{L}\left(2 + h^2\mathrm{L}\right)\|x_1 - \tilde{x}_1\| \ .
\end{aligned}
$$

Summing the two previous expression, we get the desired result for $k = 1$. □

Lemma 27. *For any $h > 0$, we have*

$$\sup_{(x,p,v)\in\mathbb{R}^{3d}} \{\|\Theta_m(x,p) - \Theta_m(x,v)\| / \|p-v\|\} \leqslant (m/h) \left\{ \left(1 + h\mathrm{L}^{1/2}\vartheta_1 (h\mathrm{L})\right)^m - 1 \right\} .$$

Proof. By Lemma 26, we have that, for any $(x,p,v) \in \mathbb{R}^{3d}$,

$$\|\operatorname{proj}_1 \circ \Lambda^{(m)}(x,p) - \operatorname{proj}_1 \circ \Lambda^{(m)}(x,v)\| \leqslant \left\{1 + h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}\right)\right\}^m \mathrm{L}^{-1/2}\|p-v\| .$$

Denote $\Lambda_1^{(i)} = \operatorname{proj}_1 \circ \Lambda^{(i)}$ and as a convention $\Lambda_1^{(0)} = \operatorname{proj}_1$. We obtain

$$\begin{aligned}
&\|\Theta_m(x,p) - \Theta_m(x,v)\| \\
&\quad\leqslant \mathrm{L}\left(\sum_{i=1}^{m-1} 2(m+1-i)\|\Lambda_1^{(i-1)}(x,p) - \Lambda_1^{(i-1)}(x,v)\| + \|\Lambda_1^{(m-1)}(x,p) - \Lambda_1^{(m-1)}(x,v)\|\right) \\
&\quad\leqslant \mathrm{L}^{1/2}\left(\sum_{i=1}^{m-1} 2(m+1-i)\left\{1 + h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}\right)\right\}^{i-1} + \left\{1 + h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}\right)\right\}^{m-1}\right)\|p-v\| \\
&\quad\leqslant 2m\mathrm{L}^{1/2}\left\{\left(1 + h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}\right)\right)^m - 1\right\} / \left(h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}\right)\right)\|p-v\| \\
&\quad\leqslant (m/h)\left\{\left(1 + h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}\right)\right)^m - 1\right\}\|p-v\| ,
\end{aligned}$$

as $\vartheta_1\left(h\mathrm{L}^{1/2}\right) \geqslant 2$. □

We can now prove Theorem 24.

Proof. Note that for any $h > 0$, $m \in \mathbb{N}^*$, we have

$$\left(1 + h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}\right)\right)^m - 1 \leqslant \exp\left\{h\mathrm{L}^{1/2}m\vartheta_1\left(h\mathrm{L}^{1/2}m\right)\right\} - 1 ,$$

as ϑ_1 is non decreasing. The function $c \to \mathrm{e}^{c\vartheta_1(c)}$ is continuous and strictly increasing, from 0 to ∞ on $\mathbb{R}$, thus $\mathrm{e}^{c\vartheta_1(c)} = 2$ admits a unique solution, for $c_0 \approx 0.29$. For $c < c_0$, we have $\mathrm{e}^{c\vartheta_1(c)} < 2$. In particular, if $h < h_0(\mathrm{L},k) = c_0/\mathrm{L}^{1/2}m$, then

$$\left\{\left(1 + h\mathrm{L}^{1/2}\vartheta_1\left(h\mathrm{L}^{1/2}k\right)\right)^m - 1\right\} < 1 .$$

We first prove that, for all $x_1 \in \mathbb{R}^d$,

$$\text{the function } p \mapsto p + h/m\Theta_m(x_1,p) \text{ is one-to-one.} \tag{41}$$

By Lemma 27, there exists $0 < \kappa < 1$ such that for all $p, v \in \mathbb{R}^d$,

$$\|H_{y_1}(p) - H_{y_1}(v)\| \leqslant \frac{h}{m}\|\Theta_m(x_1,p) - \Theta_m(x_1,v)\| \leqslant \kappa\|p-v\| ,$$

where $H_{y_1} : p \mapsto y_1 - h/m\Theta_m(x_1,p)$. Hence, by the Banach fixed point theorem, for any $y_1 \in \mathbb{R}^d$, H_{y_1} has a unique fixed point p_1 and

$$y_1 = p_1 + \frac{h}{m}\Theta_m(x_1,p_1)$$

showing (41). Hence

$$p \mapsto G_{x_1}(p) = x_1 + mhp + h^2\Theta_m(x_1, p)$$

is one-to-one. Since in addition Θ_m is continuously differentiable and the Jacobian of G_{x_1} is invertible, the function G_x is a C^1 diffeomorphism. ☐

D.1.3 Proof of (20)

The result (20) is directly linked to Theorem 28 which ensures convergence of the Markov kernel based on NICE proposals.

Theorem 28. *Assume **NICE**1 and **NICE**2. Then, the Markov kernel K defined in (39) is a π_0-reversible MH kernel with transition density*

$$q(x, y) = \varphi\big(G_x^{-1}(y)\big)\, \mathrm{J}_{G_x^{-1}}(y)\ ,$$

and acceptance probability

$$\alpha(x, y) = \mathsf{a}\left(\frac{\pi_0(y)q(y, x)}{\pi_0(x)q(x, y)}\right)\ .$$

In addition, Theorem 1 applies.

Proof. Note that for all $(x, p) \in \mathbb{R}^{2d}$,

$$\Phi^{-1} \circ \Phi(x, p) = \Phi^{-1}\big(G_x(p), H_x(p)\big) = (x, p), \tag{42}$$

where we have used $G_x(p) = \mathrm{proj}_1 \circ \Phi(x, p)$ and $H_x(p) = \mathrm{proj}_2 \circ \Phi(x, p)$. Under **NICE**2, for any $x \in \mathbb{R}^d$, $p \mapsto G_x(p)$ is a diffeomorphism. Then, plugging $y = G_x(p)$, $p = G_x^{-1}(y)$ in (42), we obtain

$$\Phi^{-1}\big(y, H_x \circ G_x^{-1}(y)\big) = \big(x, G_x^{-1}(y)\big)\ .$$

Under **NICE**1, $s \circ \Phi \circ s = \Phi^{-1}$. Then, we get

$$\Phi\big(y, -H_x \circ G_x^{-1}(y)\big) = \big(x, -G_x^{-1}(y)\big)\ .$$

Hence $G_y\big(-H_x \circ G_x^{-1}(y)\big) = x$ or equivalently, $-H_x \circ G_x^{-1}(y) = G_y^{-1}(x)$. Since φ is even, this implies

$$\varphi\big(H_x \circ G_x^{-1}(y)\big) = \varphi\big(G_y^{-1}(x)\big)\ . \tag{43}$$

Recall that $\mathrm{J}_\Phi(x, p) = 1$, for all $(x, p) \in \mathbb{R}^{2d}$. Using again $-H_x \circ G_x^{-1}(y) = G_y^{-1}(x)$ in (43), we get

$$\Phi\big(y, G_y^{-1}(x)\big) = \big(x, -G_x^{-1}(y)\big)\ .$$

Using the chain rule for Jacobian matrices, we get

$$\mathrm{J}_{G_y^{-1}}(x) = \mathrm{J}_{G_x^{-1}}(y). \tag{44}$$

Combining (43) and (44) leads to (20) by noting that

$$\frac{q(y, x)}{q(x, y)} = \frac{\varphi\big(H_x \circ G_x^{-1}(y)\big)}{\varphi\big(G_x^{-1}(y)\big)}\ .$$

Hence, the acceptance ratio α coincides with the standard MH ratio and the marginal Markov kernel K is thus π_0-reversible. We also note that $q(x, y)$ satisfies the conditions of Theorem 6 given the assumptions on φ and G_x. Moreover, K is π_0-irreducible, by Theorem 7. Then, Theorem 1 applies. ☐

D.1.4 Implementation details

Algorithm 1 presents the methodology for sampling according to the kernel K (17), which is π_0-reversible.

Algorithm 1 NICE with full refreshment at each iteration

Input: Transformation Φ and momentum-flip involution s, acceptance function a, unnormalized target density π, density φ of momentum p, initial point x_0, number of steps N
for $i = 0$ **to** $N-1$ **do**
  Draw $q_i \sim \varphi$;
  Compute proposal $(y_{i+1}, q_{i+1}) = \Phi(x_i, q_i)$;
  Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi_0(y_{i+1})\varphi(q_{i+1})}{\pi_0(x_i)\varphi(q_i)}\right) ;$$

  if $B_i \equiv 1$ **then**
    Set $x_{i+1} = y_{i+1}$;
  else
    Set $x_{i+1} = x_i$;
  end if
end for
Return $(x_{0:N})$

In order to recover persistency, as discussed in Appendix D.1.1, we consider the mixture of kernels T (40); see Algorithm 2.

Algorithm 2 NICE with randomized full refreshment

Input: Transformation Φ and momentum-flip involution s, acceptance function a, unnormalized target π, density φ of momentum p, probability of refreshment ω, initial point x_0 and initial momentum p_0, number of steps N;
for $i = 0$ **to** $N - 1$ **do**
  Draw $R_i \sim \mathrm{Ber}(\omega)$;
  if $R_i \equiv 0$ **then**
    Compute proposal $(y_{i+1}, q_{i+1}) = \Phi(x_i, p_i)$;   *### No refreshment, deterministic dynamics*
    Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi_0(y_{i+1})\varphi(q_{i+1})}{\pi_0(x_i)\varphi(q_i)}\right) \;;$$

    if $B_i \equiv 1$ **then**
      Set $(x_{i+1}, p_{i+1}) = (y_{i+1}, q_{i+1})$;   *### accept the move and keep the momentum*
    else
      Set $(x_{i+1}, p_{i+1}) = s(x_i, p_i)$;   *### reject the move and flip the momentum*
    end if
  else
    Sample $q_i \sim \varphi$;   *### Full refreshment of the momentum to update the position*
    Compute proposal $(y_{i+1}, q_{i+1}) = \Phi(x_i, q_i)$;
    Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi_0(y_{i+1})\varphi(q_{i+1})}{\pi_0(x_i)\varphi(q_i)}\right) \;;$$

    Draw $p_{i+1} \sim \varphi$;
    if $B_i \equiv 1$ **then**
      Set $x_{i+1} = y_{i+1}$;
    else
      Set $x_{i+1} = x_i$;
    end if
  end if
end for
Return $(x_{0:N})$

Algorithm 3 NICE with persistence

Input: Transformation Φ and momentum-flip involution s, acceptance function a, unnormalized target π, density φ of momentum p, hyperparameter β, initial point x_0 and initial momentum p_0, number of steps N

for $i = 0$ **to** $N-1$ **do**

Draw $u_i \sim \varphi$ and set $q_i = \beta p_i + \sqrt{1-\beta^2} u_i$;

Compute proposal $(y_{i+1}, q_{i+1}) = \Phi(x_i, q_i)$;

Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi_0(y_{i+1})\varphi(q_{i+1})}{\pi_0(x_i)\varphi(q_i)}\right) ;$$

if $B_i \equiv 1$ **then**

Set $(x_{i+1}, p_{i+1}) = (y_{i+1}, q_{i+1})$; *### accept the move and keep the momentum*

else

Set $(x_{i+1}, p_{i+1}) = s(x_i, q_i)$; *###reject the move and flip the momentum*

end if

end for

Return $(x_{0:N})$

D.2 Proof of (22)

By (45), we get

$$\begin{aligned} q\big((x,v),(y,w)\big) &= \{\rho 1_v(w) + (1-\rho)1_{-v}(w)\} q_w(x,y) \ , \\ q\big(s(y,w), s(x,v)\big) &= q\big((y,-w),(x,-v)\big) \\ &= \{\rho 1_{-w}(-v) + (1-\rho)1_w(-v)\} q_{-w}(y,x) \\ &= \{\rho 1_v(w) + (1-\rho)1_{-v}(w)\} q_{-w}(y,x) \ , \end{aligned}$$

which implies that

$$\frac{q\big(s(y,w),s(x,v)\big)}{q\big((x,v),(y,w)\big)} = \frac{q_{-w}(y,x)}{q_w(x,y)} \ .$$

The proof follows from (9).

D.3 Implementation details of Example 9

We define here a probability of refresh ω. At each iteration, we refresh the direction with probability ω, in which case we draw $v \sim \mathcal{U}\{-1,1\}$. With this definition, we can reinterpret the parameter ρ (45)

$$q\big((x,v),(y,w)\big) = \big\{\rho 1_v(w) + (1-\rho)1_{-v}(w)\big\} q_w(x,y) \ , \tag{45}$$

as $\omega = 2\rho$. In particular, we can write the lifted algorithm with randomized direction refresh in Algorithm 4.

Algorithm 4 Lifted Markov sampling

Input: Transformations $G_{1,x}, G_{-1,x}$, acceptance function a, unnormalized target π, density φ of momentum p, initial point x_0 and initial direction v_0, probability of refreshment ω, number of steps N

for $i = 0$ **to** $N-1$ **do**
  Draw $R_i \sim \mathrm{Ber}(\omega)$;
  if $R_i \equiv 1$ **then**
    Refresh direction $w_i \sim \mathcal{U}\{-1, 1\}$;
  else
    Keep direction $w_i = v_i$;
  end if
  Draw $q_i \sim \varphi$;
  Compute proposal $y_{i+1} = G_{w_i,x_i}(q_i)$;
  Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi_0(y_{i+1})\varphi\big(G^{-1}_{-w_i,y_{i+1}}(x_i)\big)\,\mathrm{J}_{G^{-1}_{-w_i,y_{i+1}}}(x_i)}{\pi_0(x_i)\varphi\big(G^{-1}_{w_i,x_i}(y_{i+1})\big)\,\mathrm{J}_{G^{-1}_{w_i,x_i}}(y_{i+1})}\right) \;;$$

  if $B_i \equiv 1$ **then**
    Set $(x_{i+1}, v_{i+1}) = (y_{i+1}, w_i)$; *### accept the move and keep the direction*
  else
    Set $(x_{i+1}, v_{i+1}) = (x_i, -w_i)$; *###reject the move and flip the direction*
  end if
end for
Return $(x_{0:N})$

D.4 Proof of Lemma 10

From (22) and (23), we get

$$\begin{aligned}\alpha\big((x,v),(y,w)\big) &= \mathsf{a}\left(\frac{q_{-w}(y,x)\pi_0(y)}{q_w(x,y)\pi_0(x)}\right)\\ &= \mathsf{a}\left(\frac{\pi_0(y)\varphi\big(\tilde G^{-1}_{-w,y}(x)\big)\,\mathrm{J}_{\tilde G^{-1}_{-w,y}}(x)}{\pi_0(x)\varphi\big(\tilde G^{-1}_{w,x}(y)\big)\,\mathrm{J}_{\tilde G^{-1}_{w,x}}(y)}\right) = \mathsf{a}\left(\frac{\mu\big(y,\tilde G^{-1}_{-w,y}(x)\big)\,\mathrm{J}_{\tilde G^{-1}_{-w,y}}(x)}{\mu\big(x,\tilde G^{-1}_{w,x}(y)\big)\,\mathrm{J}_{\tilde G^{-1}_{w,x}}(y)}\right)\;. \end{aligned} \tag{46}$$

Set $\tilde H_{w,x}(p) = \mathrm{proj}_2 \circ \Psi^w(x,p)$. Note that

$$\Psi^w(x,p) = \big(\tilde G_{w,x}(p), \tilde H_{w,x}(p)\big)\;.$$

Hence, we obtain

$$\Psi^{-w}\big(y, \tilde G^{-1}_{-w,y}(x)\big) = \big(\tilde G_{-w,y}\circ \tilde G^{-1}_{-w,y}(x), \tilde H_{-w,y}\circ \tilde G^{-1}_{-w,y}(x)\big) = \big(x, \tilde H_{-w,y}\circ \tilde G^{-1}_{-w,y}(x)\big)\;,$$

which implies,

$$\begin{aligned}\big(y, \tilde{G}_{-w,y}^{-1}(x)\big) &= \Psi^{w}\big(x, \tilde{H}_{-w,y} \circ \tilde{G}_{-w,y}^{-1}(x)\big) \;, \\ &= \big(\tilde{G}_{w,x} \circ \tilde{H}_{-w,y} \circ \tilde{G}_{-w,y}^{-1}(x), \tilde{H}_{w,x} \circ \tilde{H}_{-w,y} \circ \tilde{G}_{-w,y}^{-1}(x)\big) \;.\end{aligned} \tag{47}$$

This identity in particular shows that $y = \tilde{G}_{w,x} \circ \tilde{H}_{-w,y} \circ \tilde{G}_{-w,y}^{-1}(x)$ or equivalently $\tilde{G}_{w,x}^{-1}(y) = \tilde{H}_{-w,y} \circ \tilde{G}_{-w,y}^{-1}(x)$, which used in (47) establishes

$$\big(y, \tilde{G}_{-w,y}^{-1}(x)\big) = \Psi^{w}\big(x, \tilde{G}_{w,x}^{-1}(y)\big) \;. \tag{48}$$

Set $A_w(x,y) = \big(y, \tilde{G}_{-w,y}^{-1}(x)\big)$ and $B_w(x,y) = \big(x, \tilde{G}_{w,x}^{-1}(y)\big)$. Note that $\mathrm{J}_{A_w}(x,y) = \mathrm{J}_{\tilde{G}_{-w,y}^{-1}}(x)$, $\mathrm{J}_{B_w}(x,y) = \mathrm{J}_{\tilde{G}_{w,x}^{-1}}(y)$ and by (48) and the chain rule

$$\mathrm{J}_{A_w}(x,y) = \mathrm{J}_{\Psi^w}\big(B_w(x,y)\big)\,\mathrm{J}_{B_w}(x,y) \;,$$

which implies

$$\mathrm{J}_{\Psi^w}\big(x, \tilde{G}_{w,x}^{-1}(y)\big) = \frac{\mathrm{J}_{\tilde{G}_{-w,y}^{-1}}(x)}{\mathrm{J}_{\tilde{G}_{w,x}^{-1}}(y)}. \tag{49}$$

The proof of Lemma 10 is concluded by plugging (48) and (49) into (46).

D.5 Lifted acceptance probability with deterministic proposals

In this case $\Phi(x,p,v) = (\Psi^v(x,p), v)$, $(x,p) \in \mathbb{R}^{2d}$, $s \in \mathsf{V}$. Clearly, $\Phi^{-1}(x,p,v) = (\Psi^{-v}(x,p), v)$ and it is easily checked that $\Phi^{-1} = s \circ \Phi \circ s$. Denote

$$\pi\big(\mathrm{d}(x,p,v)\big) = \pi_0(x)\varphi(p)\mathrm{d}x\mathrm{d}p\rho(\mathrm{d}v) = \mu(x,p)\mathrm{d}x\mathrm{d}p\rho(\mathrm{d}v) \;.$$

To compute the acceptance probability (12), we need to evaluate the density $k(z) = \mathrm{d}\mu/\mathrm{d}\lambda(z)$, where

$$\lambda = \pi + \Phi_{\#}^{-1}\pi.$$

Let $f\colon \mathbb{R}^d \times \mathbb{R}^d \times \mathsf{V} \to \mathbb{R}_+$ be a measurable function. We get

$$\begin{aligned}\lambda(f) &= \int f(x,p,v)\mu(x,p)\mathrm{d}x\mathrm{d}p\rho(\mathrm{d}v) + \int f(\Psi^{-v}(x,p),v)\mu(x,p)\mathrm{d}x\mathrm{d}p\rho(\mathrm{d}v) \\ &= \int f(x,p,v)\mu(x,p)\mathrm{d}x\mathrm{d}p\rho(\mathrm{d}v) + \int f(x',p',v)\mu\big(\Psi^{v}(x',p')\big)\,\mathrm{J}_{\Psi^v}(x',p')\mathrm{d}x'\mathrm{d}p'\rho(\mathrm{d}v) \;.\end{aligned}$$

Therefore, we get

$$\frac{\mathrm{d}\lambda}{\mathrm{dLeb}_{2d} \otimes \rho}(x,p,v) = \mu(x,p) + \mu\big(\Psi^v(x,p)\big)\,\mathrm{J}_{\Psi^v}(x,p) \;.$$

This implies that, for all $(x,v) \in \mathbb{R}^d \times \mathsf{V}$,

$$k(x,p,v) = \frac{\mu(x,p)}{\mu(x,p) + \mu\big(\Psi^v(x,p)\big)\,\mathrm{J}_{\Psi^v}(x,p)} \;.$$

Since $s \circ \Phi(x,p,v) = (\Psi^v(x,p), -v)$, we obtain

$$
\begin{aligned}
k\big(s \circ \Phi(x,p,v)\big) &= \frac{\mu\big(\Psi^v(x,p)\big)}{\mu\big(\Psi^v(x,p)\big) + \mu(x,p)\,\mathrm{J}_{\Psi^{-v}}\big(\Psi^v(x,p)\big)} \\
&= \frac{\mu\big(\Psi^v(x,p)\big)\,\mathrm{J}_{\Psi^v}(x,p)}{\mu(x,p) + \mu\big(\Psi^v(x,p)\big)\,\mathrm{J}_{\Psi^v}(x,p)} \;,
\end{aligned}
$$

where we have used $\mathrm{J}_{\Psi^v}(x,p) = 1/\,\mathrm{J}_{\Psi^{-v}}\big(\Psi^v(x,p)\big)$. Therefore, the acceptance probability is given by

$$
\alpha\big((x,v),(y,w)\big) = \begin{cases} \mathsf{a}\left(\frac{\mu\left(\Psi^v(x,p)\right)\mathrm{J}_{\Psi^v}(x,p)}{\mu(x,p)}\right) \;, \text{ if } \mu(x,p) > 0, (y,w) = (\Psi^v(x,p), v), \\ 1, \;\text{ if } \mu(x,p) = 0 \;\text{ or }\; (y,w) \neq (\Psi^v(x,p), v) \;. \end{cases}
$$

D.6 L2HMC Algorithms

In this section, we discuss the sampling algorithms associated to the L2HMC kernel, Example 11. Again, we first describe a version of this algorithm in which the momentum is fully refreshed at each iteration. This is a lifted version of the original L2HMC algorithm Levy et al. (2017), because we keep the direction variable at each iteration instead of refreshing the direction at each iteration. Consider the following assumption.

L2HMC1. *For all $v, x \in \{-1,+1\} \times \mathbb{R}^d$,*

$$
G_{v,x} \colon p \mapsto \mathrm{proj}_1 \circ \Psi^v(x,p)
$$

is a C^1-diffeomorphism.

As in the NICE case, establishing **L2HMC**1 requires conditions on the mapping Ψ defining the L2MHC transitions, which is subject to a future work. Under **L2HMC**1, Lemma 10 shows that the lifted L2HMC with full momentum refresh satisfies the assumption of Proposition 8. We may therefore apply Theorem 6 to show convergence of the algorithm in the sense of Theorem 1.

As said above, the original L2HMC algorithm (Algorithm 5) refreshes at each iteration both the direction and the momentum, whereas the lifted algorithm keeps the direction and refreshes only the momentum. Similar to the NICE case, define the marginal kernel, acting on the position only:

$$
K(x, \mathrm{d}y) = K_\alpha(x, \mathrm{d}y) + \big(1 - \bar{\alpha}(x)\big)\delta_x(\mathrm{d}y) \;,
$$

where $\bar{\alpha}(x) = K_\alpha(x, \mathbb{R}^d)$ and for a measurable nonnegative function f,

$$
K_\alpha f(x) = \iint f\big(\mathrm{proj}_1 \circ \Psi^v(x,p)\big)\bar{\alpha}(x,p,v) Q\big((x,p,v), \mathrm{d}(y,q,w)\big)\varphi(p)\mathrm{d}p\rho(\mathrm{d}v) \;,
$$

where

$$
Q\big((x,p,v), \mathrm{d}(y,q,w)\big) = \delta_{\Psi^v(x,p)}\big(\mathrm{d}(y,q)\big)\delta_v(\mathrm{d}w) \;.
$$

and

$$
\bar{\alpha}(x,p,v) = \mathsf{a}\left(\mu\big(\Psi^v(x,p)\big)/\mu(x,p)\;\; \mathrm{J}_{\Psi^v}(x,p)\right) \;,
$$

Algorithm 5 Original L2HMC

Input: Transformation Ψ, acceptance function a, unnormalized target π, density φ of momentum p, initial point x_0, number of steps N

for $i = 0$ **to** $N-1$ **do**

  Refresh momentum $q_i \sim \varphi$ and direction $v_i \sim \mathcal{U}\{-1, 1\}$;

  Compute proposal $(y_{i+1}, q_{i+1}) = \Psi^{v_i}(x_i, q_i)$;

  Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi(y_{i+1})\varphi(q_{i+1})}{\pi(x_i)\varphi(q_i)} \mathrm{J}_{\Psi^{v_i}}(x_i, q_i)\right) ;$$

  if $B_i \equiv 1$ **then**

    Set $x_{i+1} = y_{i+1}$;

  else

    Set $x_{i+1} = x_i$;

  end if

end for

Return $(x_{0:N})$

Recall that the Markov kernel $Q_\alpha\big(x, p, v; \mathrm{d}(y, q, w)\big) = \bar{\alpha}(x, p, v) Q\big((x, p, v), \mathrm{d}(y, q, w)\big)$ is (π, S)-reversible. Let f and g be two positive measurable functions on $\mathbb{R}^d$. Since Q_α is (π, S)-reversible and ρ is symmetric, we get

$$\begin{aligned}
\int \pi_0(\mathrm{d}x) K_\alpha f(x) g(x) &= \int \pi\big(\mathrm{d}(x, p, v)\big) Q_\alpha\big((x, p, v); \mathrm{d}(y, q, w)\big) f(y) g(x) \\
&= \int \pi\big(\mathrm{d}(y, q, w)\big) S Q_\alpha S g(y) f(y) \\
&=^{(3)} \int \pi\big(\mathrm{d}(y, q, w)\big) \bar{\alpha}(y, q, -w) Q\big((y, q, -w); \mathrm{d}(x, p, v)\big) g(x) f(y) \\
&=^{(4)} \int \pi\big(\mathrm{d}(y, q, w)\big) \bar{\alpha}(y, q, w) Q\big((y, q, w); \mathrm{d}(x, p, v)\big) g(x) f(y) \\
&=^{(5)} \int \pi_0(\mathrm{d}x) K_\alpha g(y) f(y)
\end{aligned}$$

where we have used in (3) that $S Q_\alpha S g(y, q, w) = \bar{\alpha}(y, q, -w) \int Q\big((y, q, -w); \mathrm{d}(x, p, v)\big) g(x)$, in (4) the symmetry of ρ and finally in (5) the definition of K_α. Hence the L2HMC kernel is π_0-reversible. Moreover, note that under **L2HMC**1, we obtain

$$K_\alpha f(x) = \int \left\{ \int \bar{\alpha}(x, G_{v,x}^{-1}(y), v) \varphi\big(G_{v,x}^{-1}(y)\big) \mathrm{J}_{G_{v,x}^{-1}}(y) \rho(\mathrm{d}v) \right\} f(y) \mathrm{d}y \, .$$

Denoting by $q_v(x, y)$ the transition density $q_v(x, y) = \varphi\big(G_{v,x}^{-1}(y)\big) \mathrm{J}_{G_{v,x}^{-1}}(y)$ and then setting $q(x, y) = \int q_v(x, y) \rho(\mathrm{d}v)$, we finally get

$$K_\alpha f(x) = \int \alpha(x, y) q(x, y) f(y) \mathrm{d}y$$

Algorithm 6 Lifted L2HMC with full momentum refreshment

Input: Transformation Ψ, acceptance function a, unnormalized target π, density φ of momentum p, probability of direction refreshment ω, initial point x_0 and initial direction v_0, number of steps N

for $i = 0$ **to** $N-1$ **do**
  Draw $R_i \sim \mathrm{Ber}(\omega)$;
  if $R_i \equiv 1$ **then**
    Refresh direction $w_i \sim \mathcal{U}\{-1, 1\}$;
  else
    Keep direction $w_i = v_i$;
  end if
  Refresh momentum $q_i \sim \varphi$;
  Compute proposal $(y_{i+1}, q_{i+1}) = \Psi^{w_i}(x_i, q_i)$;
  Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi(y_{i+1})\varphi(q_{i+1})}{\pi(x_i)\varphi(q_i)} \, \mathrm{J}_{\Psi^{w_i}}(x_i, q_i)\right) \; ;$$

  if $B_i \equiv 1$ **then**
    Set $(x_{i+1}, v_{i+1}) = (y_{i+1}, w_i)$; *### accept the move and keep the direction*
  else
    Set $(x_{i+1}, v_{i+1}) = (x_i, -w_i)$; *###reject the move and flip the direction*
  end if
end for
Return $(x_{0:N})$

with

$$\alpha(x, y) = \frac{\int \bar{\alpha}(x, G_{v,x}^{-1}(y), v)\varphi\big(G_{v,x}^{-1}(y)\big) \, \mathrm{J}_{G_{v,x}^{-1}}(y)\rho(\mathrm{d}v)}{\int \varphi\big(G_{v,x}^{-1}(y)\big) \, \mathrm{J}_{G_{v,x}^{-1}}(y)\rho(\mathrm{d}v)}$$

Write now, following (Tierney, 1994),

$$r(x, y) = \frac{\pi_0(x)q(x, y)}{\pi_0(y)q(y, x)} \; .$$

Moreover, by Lemma 10, we can write

$$\bar{\alpha}(x, G_{v,x}^{-1}(y), v) = \mathsf{a}\left(\frac{\pi_0(y)q_{-v}(y, x)}{\pi_0(x)q_v(x, y)}\right) \; .$$

In that case, we have

$$\begin{aligned}
\alpha(x,y)\pi_0(x)q(x,y) &= \int \pi_0(x)q_v(x,y)\mathsf{a}\left(\frac{\pi_0(y)q_{-v}(y,x)}{\pi_0(x)q_v(x,y)}\right)\rho(\mathrm{d}v) \\
&= \int \pi_0(y)q_{-v}(y,x)\mathsf{a}\left(\frac{\pi_0(x)q_v(x,y)}{\pi_0(y)q_{-v}(y,x)}\right)\rho(\mathrm{d}v) \\
&= \int \pi_0(y)q_w(y,x)\mathsf{a}\left(\frac{\pi_0(x)q_{-w}(x,y)}{\pi_0(y)q_w(y,x)}\right)\rho(\mathrm{d}w)\ ,
\end{aligned}$$

where we have use the fact that $t\mathsf{a}(1/t) = \mathsf{a}(t)$ and the change of variable $w = -v$. Then,

$$\alpha(x,y)\pi_0(x)q(x,y) = \alpha(y,x)\pi_0(y)q(y,x)\ .$$

We thus have $\alpha(x,y)r(x,y) = \alpha(y,x)$, which proves by (Tierney, 1994, Theorem 2) that the ratio α is exactly the classical MH ratio satisfying the detailed balance condition.

To retrieve persistency, we can use as for the NICE algorithm a mixture of a deterministic L2HMC move and a full independent refreshment of the position, momentum and the direction; see Algorithm 7.

Algorithm 7 Lifted L2HMC with randomized full refreshment

Input: Transformation Ψ, acceptance function a, unnormalized target π, density φ of momentum p, probability of refreshment ω, initial point x_0, initial momentum p_0 and initial direction v_0, number of steps N

for $i = 0$ **to** $N - 1$ **do**
  Draw $R_i \sim \mathrm{Ber}(\omega)$;
  if $R_i \equiv 0$ **then**
    Compute proposal $(y_{i+1}, q_{i+1}) = \Psi^{v_i}(x_i, p_i)$; *###No refreshment, deterministic dynamics*
    Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi(y_{i+1})\varphi(q_{i+1})}{\pi(x_i)\varphi(p_i)}\,\mathrm{J}_{\Psi^{v_i}}(x_i, p_i)\right) \;;$$

    if $B_i \equiv 1$ **then**
      Set $(x_{i+1}, p_{i+1}, v_{i+1}) = (y_{i+1}, q_{i+1}, v_i)$; *### accept the move and keep the direction*
    else
      Set $(x_{i+1}, p_{i+1}, v_{i+1}) = (x_i, p_i, -v_i)$; *###reject the move and flip the direction*
    end if
  else
    Draw $q_i \sim \varphi$, $w_i \sim \mathcal{U}\{-1, 1\}$; *### refresh independently the momentum and the direction*
    Compute proposal $(y_{i+1}, q_{i+1}) = \Psi^{w_i}(x_i, q_i)$;
    Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi(y_{i+1})\varphi(q_{i+1})}{\pi(x_i)\varphi(q_i)}\,\mathrm{J}_{\Psi^{w_i}}(x_i, q_i)\right) \;;$$

    Draw $p_{i+1} \sim \varphi$, $v_{i+1} \sim \mathcal{U}\{-1, 1\}$; *### refresh the momentum and the direction*
    if $B_i \equiv 1$ **then**
      Set $x_{i+1} = y_{i+1}$;
    else
      Set $x_{i+1} = x_i$;
    end if
  end if
end for
Return $(x_{0:N})$

Much like the persistent HMC algorithm, we may also design a lifted persistent HMC algorithm in which, at each iteration, we keep the direction and partially refresh the momentum using an autoregressive scheme; see Algorithm 8.

Algorithm 8 Lifted L2HMC with persistence

Input: Transformation Ψ, acceptance function a, unnormalized target π, density φ of momentum p, hyperparameter β, initial point x_0, initial momentum p_0 and initial direction v_0, number of steps N

for $i = 0$ **to** $N-1$ **do**

Sample $u_i \sim \varphi$ and refresh momentum $q_i = \beta p_i + \sqrt{1-\beta^2} u_i$; *### partially update the momentum*

Compute proposal $(y_{i+1}, q_{i+1}) = \Psi^{v_i}(x_i, q_i)$;

Draw $B_i \sim \mathrm{Ber}(a_i)$ where

$$a_i = \mathsf{a}\left(\frac{\pi(y_{i+1})\varphi(q_{i+1})}{\pi(x_i)\varphi(q_i)} \, \mathrm{J}_{\Psi^{v_i}}(x_i, q_i)\right) \; ;$$

if $B_i \equiv 1$ **then**

Set $(x_{i+1}, p_{i+1}, v_{i+1}) = (y_{i+1}, q_{i+1}, v_i)$; *### accept the move and keep the direction*

else

Set $(x_{i+1}, p_{i+1}, v_{i+1}) = (x_i, p_i, -v_i)$; *### reject the move and flip the direction*

end if

end for

Return $(x_{0:N})$